

Genetics without borders

A UK government scheme to establish nationality through DNA testing is scientifically flawed, ethically dubious and potentially damaging to science.

Until a few years ago, the genetic variation of humans was understood only in terms of superficial characteristics, such as hair and skin colour. Today, thanks to the advent of cheap, fast genetic sequencing and DNA-microarray technologies, population geneticists can chart such variations in a more systematic way. Yet most experts agree that these studies are still in their infancy.

So it was with understandable incredulity that researchers received a plan by the UK Border Agency to use genetics to determine nationality — specifically, the origin of asylum-seekers claiming to be from war-torn Somalia. The agency's pilot programme, which began last month, aims to determine whether some 100 individuals really are Somali nationals by checking them for the individual DNA variants known as single nucleotide polymorphisms (SNPs) in mitochondrial DNA, on the Y chromosome and elsewhere in the genome. The scheme will also use isotopic ratios of elements found in hair and fingernails — which can vary depending on a person's diet or environment — to try to establish where the migrants previously lived.

The border agency says that the project has undergone scientific peer review, although it is difficult to say by whom: several geneticists contacted by *Nature* saw a preliminary proposal from the UK government in 2007, and warned that it was unlikely to work.

It is true that the recent development of large SNP databases have made it possible to determine the geographic origins of Europeans to within a few hundred kilometres (see *Nature* 456, 98–101; 2008). But comparable data on many human populations, especially in regions such as Africa, remain patchy at best, and it is unclear what data the border agency will use to establish the origins of these particular asylum-seekers.

On a more fundamental level, the idea that genetic variability follows man-made national boundaries is absurd. Cross-border migration

is common throughout the world; Y-chromosome analysis can easily be thrown off by a distant male ancestor; and SNP-based identifications are inexact to say the least. As an example of this last point, individuals whose parents come from two geographic regions are often classed into a third region from which neither parent originated.

The use of isotopic analysis for identifying nationality is also unproven. Although it may be possible to use isotopic ratios to determine the region in which a person has recently lived, it cannot provide definitive evidence of their origins.

These problems seem to be ignored in the guidelines provided to border agents testing the asylum-seekers. Given the scientific credibility of DNA evidence, it is not difficult to imagine that these agents — who are presumably not geneticists — might place undue weight on results that are, at best, difficult to interpret and, at worst, spurious.

Migration organizations and geneticists alike have been vocal in their protests against the plan, and in response the UK government seems to have backpedalled. In a statement released earlier this week from the Home Office, which runs the border agency, the programme was described as only a proof-of-concept project that would not be used to make decisions about any asylum-seeker. But the government should cancel this scientifically dubious and politically sensitive programme outright. If it is allowed to continue, it could easily lead to a public backlash in the very populations that geneticists need to study to understand human origins and the genetic underpinnings of disease. Geneticists, and indeed all scientists, should decry the plan and make it clear that the science does not support it. ■

"The idea that genetic variability follows national boundaries is absurd."

Putting DNA to the test

Genetic-testing companies lack regulation, and a list of guiding principles does not go far enough.

The availability of affordable, direct-to-consumer genetic tests has mushroomed, leaving regulation lagging behind. Dozens of companies now offer inexpensive home kits that allow people to spit into tubes, send the samples for DNA analysis and receive a report that allegedly details their ancestry or their possible susceptibility to a long list of disorders that have been linked — often tenuously — to particular genes. But the value of these tests remains debatable, which is why the industry needs a strong set of quality standards and codes of conduct to protect both its consumers and its own credibility.

The UK Human Genetics Commission (HGC) took a welcome step in that direction last month when it issued a set of principles to help guide consumers and to promote high standards and consistency among personal-genomics providers. But the HGC's guiding principles — which are under public review until early December — focus largely on reining in companies' promotional messages so that they reflect the limited utility of genetic testing, and to make would-be customers more aware of what they can realistically expect to learn from the tests. Most DNA testing companies say they are already doing just that, emphasizing that what they provide is information, not medical diagnoses.

The question is what happens if or when prices drop further and the tests become more popular. They are already being marketed over the Internet with little oversight, and it seems likely that increasing numbers of people will be turning to personal-genomics companies in search of definitive answers about how to improve or safeguard

their health. But the available answers are rapidly becoming less definitive: the ever-increasing number of genome-wide association studies, which provide a major portion of the genetic markers for disease risks, offer results that even researchers can find hard to interpret, and often flag up markers that are not the most useful predictors of complex traits (see page 712).

This means that customers will frequently receive results telling them only that they face the ambiguous possibility of a somewhat elevated risk of a little-understood disorder. Presumably, most customers in that position will simply be more vigilant about disease screening. But if the ambiguous, slightly elevated risk relates to a frightening condition such as breast cancer, some individuals might feel compelled to undertake drastic and perhaps needless measures, such as prophylactic mastectomy.

The HGC's principles, if adopted, should help to minimize such panic reactions. For example, the HGC advocates that genetic counselling be provided both before and after testing for serious hereditary diseases. But there is room to go further and, on page 724, Craig

Venter and his colleagues offer nine recommendations for how to do so. For example, Venter's team urges companies to agree on a core set of non-ambiguous genetic markers — ones that put a carrier at high risk of developing a specific condition. Companies would be well-advised to follow this recommendation, as such an agreement would help to avoid conflicting messages and inconsistent results across the industry.

Ultimately, however, government regulators may feel compelled to step in on the grounds that industry self-policing no longer offers consumers adequate protection. The US Food and Drug Administration has already recruited bioethicist Alta Charo, in part to advise commissioner Peggy Hamburg on a comprehensive approach to regulating these tests.

Government regulators should proceed with care, given the dizzying speed at which the science of personalized genomics is advancing. But in the interim, DNA-test providers should up their game by providing only clinically useful information and spelling out exactly how much biology remains unknown. ■

How to win trust over flu

Mass-vaccination campaigns for the pandemic H1N1 virus must take public concerns into account.

As countries roll out their campaigns for large-scale vaccination against pandemic H1N1 flu, a poll released last week by the Harvard School of Public Health in Boston, Massachusetts, found that only four in ten US adults have definitely decided to get vaccinated themselves, and just half plan to get the shot for their children (go.nature.com/wiB8V3). Harvard's results parallel those from other surveys, both inside and outside the United States, all of which suggest that many people are still dubious about the vaccine. Public-health authorities, who are keen to contain the pandemic's spread, need to realize that their best hope of dealing with such public reluctance is to patiently address the concerns that underlie it.

Sometimes, it's true, those concerns go beyond any appeal to reason. They grow out of a visceral mistrust of authority in general — and of government, regulatory agencies, medical researchers and multinational pharmaceutical companies, in particular. A sophisticated anti-vaccine movement has emerged that plays on this wariness, and helps to feed the conspiracy theories about the H1N1 vaccine that are circulating on the Internet and in viral e-mails.

But far more often, say researchers who have studied this subject, people are assessing vaccination through a perfectly rational cost-benefit analysis. There is a widespread public perception, for example, that the vaccine's safety trials have been rushed — the Harvard study found that possible side effects were respondents' main concern — and that H1N1 flu is mild. As a result, many feel no urgent need to be vaccinated, preferring to hold off until they see how the first phases of the vaccination programme go. Indeed, the Harvard poll also found that some 60% of those who don't intend to get a shot are open to changing their mind if people in their community become severely ill or die.

Such deliberations reflect a perfectly legitimate decision-making

process, says Peter Sandman, a risk-communication consultant in Princeton, New Jersey. And governments, he advises, should frame their public-education campaigns in ways that respect people's judgement and their wait-and-see attitude. Research in risk communication strongly shows that when over-eager officials pressure members of the public who are already sceptical and ambivalent, while being openly dismissive of public concerns, they only end up stoking resistance.

Instead, officials should focus on providing people with the information they need to make good choices for themselves. This should include reminders that coincidences do happen: in any mass-vaccination campaign, at least a few people will fall ill immediately after receiving their shot for reasons that have nothing to do with the vaccine — a possibility vividly highlighted last week by the death of a 14-year-old British girl hours after receiving a vaccine against human papilloma virus. Regulatory authorities need to better explain the extensive safety tests that vaccines undergo and, at the same time, build confidence by being utterly transparent in the reporting and investigation of any suspect events linked to vaccination.

The public-education campaign should also correct the misconception that H1N1 flu is mild. It is mild in most who catch it. But for those individuals — mainly young adults — who will develop the severe form, H1N1 is life-threatening. Moreover, because the virus is new and immunity is lacking, many more people will get it than is typical for seasonal flu, and the toll of serious illness and deaths will accordingly be that much higher.

Finally, people should be reminded that vaccination isn't just about protecting themselves; it's also about not spreading the flu to others, which, importantly, alleviates pressure on overstretched hospitals. Campaigns should give altruistic appeals far more prominence than they typically have in the past; research shows that they can be surprisingly effective.

More generally, for officials and researchers at all levels, the scepticism over the pandemic vaccine should serve as a timely reminder of the imperative to work to gain greater public trust in science-based advice and in those who give it. ■

RESEARCH HIGHLIGHTS

CANCER BIOLOGY

Stem cell-cancer link

Nature Genet. doi:10.1038/ng.465 (2009)

The *SOX2* gene, famous for its role in helping to reprogram adult cells into stem cells, is also a cancer driver.

Matthew Meyerson of the Dana-Farber Cancer Institute in Boston, Massachusetts, and his colleagues searched genome-wide for tumour-promoting genes in human samples of lung and oesophageal squamous-cell carcinomas. They found that a region around *SOX2* was frequently replicated in both diseases. *SOX2* expression is necessary for the growth of lung and oesophageal squamous-cell cancer lines. Overactivating *SOX2* also turned normal cells cancerous with help from a couple of other genes.

ECOLOGY

Wildebeest chain reaction

PLoS Biol. 7, e1000210 (2009)

One change in an ecosystem can have far-reaching effects. This is evident in the Serengeti in East Africa, where tree density has increased since the 1960s, when the rinderpest virus, which attacks wildebeest, was eradicated. To figure out what the connection between these events might be, Ricardo Holdo of the University of Florida in Gainesville and his colleagues compared ten models of tree and fire dynamics on the famous savannah.

The researchers conclude that after the disease was wiped out, wildebeest grew in number and ate more grasses. With less grass to burn, fires decreased in frequency and more seedlings were able to grow to maturity. Other factors such as climate change and browsing by elephants seemed to have less of a role. The team adds that this shift means that the Serengeti may have become a carbon sink.



Boys against girls

Science doi:10.1126/science.1174705 (2009)

Male cichlid fishes in East Africa's Lake Malawi have evolved striking coloration (top right) to compete for females. The objects of their affections, meanwhile, tend to sport inconspicuous brown scales (top left).

An exception is the

'orange-blotch' trait, which is found almost exclusively in females (bottom left) and provides them with camouflage. When it does occur in males, it disrupts their patterning (bottom right), reducing their fitness.

Thomas Kocher and his colleagues at the University of Maryland in College Park have found that this trait is caused by a mutation in the

Pax7 gene, which is tightly linked to a new female-sex-determining gene. The close linkage between the mutated gene and the female sex determiner ensures that orange-blotch is expressed mainly in females. Such sexual conflicts can lead to the evolution of new sex-determining systems and many other traits, the team suggests.

GEOSCIENCE

Earth's magnetic personality

Geochim. Geophys. Geosys.

doi:10.1029/2009GC002496 (2009)

When did our planet develop the roiling convection pattern that churns its metallic core and gives rise to its magnetic field?

Ancient rocks from South Africa's Barberton greenstone belt reveal that such currents in the core must have started by 3.45 billion years ago, significantly earlier than had been established from previous rock evidence, report John Tarduno of the University of Rochester in New York and his colleagues. The rocks sport a magnetic signature indicating that the planet had developed a substantial magnetic field by that time.

Past work by Tarduno and his co-workers had provided evidence that a significant magnetic field was present by 3.2 billion years ago.

ANALYTICAL CHEMISTRY

Gloop monitor

Angew. Chem. Int. Edn doi:10.1002/anie.200902360 (2009)

Mass spectrometry is an invaluable tool for analysing substances' molecular compositions. But using it on

viscous liquids such as toothpaste has required the time-consuming step of taking selective extracts from the liquids. Now these sticky complex mixtures can be monitored directly in a flask.

Renato Zenobi at the Swiss Federal Institute of Technology in Zurich and Huanwen Chen at the East China Institute of Technology in Fuzhou and their colleagues blew nitrogen gas through samples of toothpaste, honey and olive oil to create bubbles that carry molecules up to the sample's surface. There, the bubbles burst, creating aerosols that can be analysed in a standard mass spectrometer.

The team used the technique to track the progress of chemical reactions in viscous ionic liquids, which are increasingly popular solvents.

AGEING

Live longer, but how?

Science 326, 140-144 (2009)

Caloric restriction extends the lifespan of many model organisms, a finding that has prompted some people to drastically reduce their food intake in the hope of upping their longevity. But how exactly caloric restriction staves off death remains unknown.

Dominic Withers of University College London and his colleagues found that mice in which the gene *S6k1* was deleted lived



for about 80 days — or 9% — longer than control mice, with females surviving 20% longer. The animals were also less likely to develop certain signs of ageing, such as loss of insulin sensitivity. The gene-expression patterns of the mice were similar to those in mice undergoing long-term caloric restriction, suggesting that manipulating S6K1 signalling could be a strategy for researchers seeking drugs that mimic the positive effects of this regime.

STEM-CELL BIOLOGY

Rebooting cord blood cells

Cell Stem Cell 5, 434–441; 353–357 (2009)

A well-known cocktail of genes can reset many adult cells to 'pluripotency', a state from which they can develop into almost any tissue. Now, two groups have derived induced pluripotent stem (iPS) cell lines from umbilical cord blood, a source that could be clinically useful.

Ulrich Martin of Hannover Medical School in Germany and his colleagues created the cells from cord blood using four genes, *OCT4*, *SOX2*, *NANOG* and *LIN28*. Juan Carlos Izpisua Belmonte at the Salk Institute for Biological Studies in La Jolla, California, and his collaborators generated iPS cells using as few as two genes, *OCT4* and *SOX2*.

Cord-blood cells have not acquired as many mutations as other cells, so stem cells such as these might be less prone to turn cancerous if used as therapy. Nevertheless, both groups used viruses to insert the genes, reducing the cells' direct therapeutic utility.

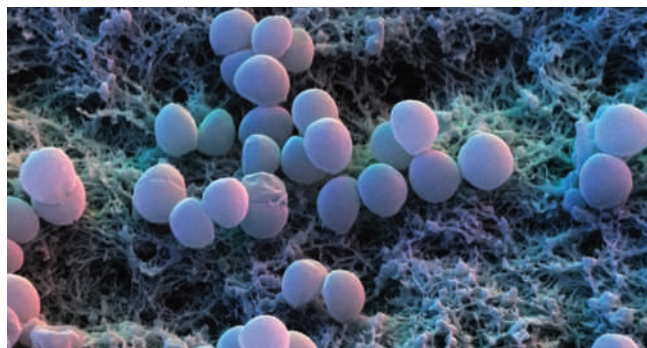
MICROBIOLOGY

Bacteria fight back

J. Exp. Med. doi:10.1084/jem.20090097 (2009)

A bacterium that causes many skin and bloodborne infections and the bacterium responsible for anthrax both synthesize the same molecule to evade host immune responses.

Olaf Schneewind and his colleagues at the University of Chicago in Illinois examined the ability of *Staphylococcus aureus* (pictured below) to survive in rodent blood. They found that an enzyme anchored to the bacterium's cell wall produces adenosine, a key signalling molecule, during infection to protect the bacteria from attack by white blood cells. The anthrax pathogen *Bacillus anthracis* uses the same mechanism for survival.



CHEMISTRY

Microwave magic

Angew. Chem. Int. Edn doi:10.1002/anie.200904185 (2009)

Microwave irradiation is commonly used to boost the speed and yield of chemical reactions, but how it works has been unclear. Debate centres on whether it is

due to the heat supplied or to some effect of microwaves' electromagnetic field on the reaction components.

Oliver Kappe and his colleagues at Karl Franzens University in Graz, Austria, have separated the two effects using silicon carbide vials, which transmit the heat but block out their electromagnetic field.

When they measured reaction time and product yield in 18 microwave-enhanced reactions, the researchers obtained almost identical results with silicon carbide vials as with Pyrex containers. This suggests that in most cases heat is responsible for the benefits of microwave chemistry.

MATERIALS SCIENCE

No gas from glass

Nature Mater. doi:10.1038/nmat2542 (2009)

Where temporary surgical implants are concerned, materials that decompose safely over time eliminate the need for costly and painful removal. Magnesium in its crystalline form has been used in some devices because it is about as strong as bone. But when it corrodes, it releases hydrogen gas, raising the risk of gas pockets being formed in tissues.

Jörg Löffler and his colleagues at the Swiss Federal Institute of Technology in Zurich sharply reduced gas release by creating glassy magnesium–zinc alloys. The gas reduction — which took place only if the alloy's zinc content was at least 28% — happened because a dense outer layer of zinc oxide or carbonate forms when the magnesium in these alloys decomposes, preventing hydrogen from forming bubbles.

JOURNAL CLUB

Judith E. Mank
Edward Grey Institute,
Department of Zoology,
University of Oxford, UK

An evolutionary biologist compares genomic complexity to modern art.

Like many students of evolutionary biology, I was taught that genes encode physical traits, or 'phenotypes', that are the focus of natural selection — a model with clear, direct links and few, if any, complications. Over the past few years, I have found it

increasingly difficult to reconcile this simple model connecting genes and the organisms they encode with the burgeoning data of systems biology, which show the genome as a heaving tangle of interconnections. Given the complexity of the genome, how can selection target any single gene without unintended consequences?

Trudy Mackay at North Carolina State University in Raleigh and her collaborators have begun to resolve the opposing genomic and evolutionary world views by examining the systems genetics that underlie phenotypes in the fruitfly *Drosophila melanogaster*

(J. F. Ayroles *et al.* *Nature Genet.* 41, 299–307; 2009). They do this by comparing data on the abundance of more than 10,000 DNA transcripts with whole-organism traits, such as fitness and lifespan, in 40 fruitfly lines.

The researchers show that aggregates of genes correlate with distinct characteristics in flies, and that these modules are connected, with groups of genes associated with multiple phenotypic traits. This elegant complexity is best conveyed by the figures in the paper, some of which look as though they were lifted off the walls of a modern-art gallery.

The group's work provides a post-genomic framework for dissecting the intricate underpinnings of organismal biology. More importantly, the paper demonstrates that key topics in traditional evolutionary studies, such as heritability, and more recent concepts, such as pleiotropy (whereby one gene affects multiple traits), are related. As such, they must be considered together to build a complete understanding of how selection acts through the phenotype to sculpt the genome.

Discuss this paper at <http://blogs.nature.com/nature/journalclub>

NEWS BRIEFING

● POLICY

Climate law: Democrats in the US Senate unveiled their climate legislation on 30 September. Based largely on the Waxman–Markey bill passed by the House of Representatives in June, it proposes a cap-and-trade system to reduce greenhouse-gas emissions by 20% by 2020 and 83% by 2050, compared with 2005 levels. An initial committee vote could come later this month. On the same day, the US Environmental Protection Agency proposed a rule that would require major industrial facilities to use “best available” technologies to reduce greenhouse-gas emissions. See go.nature.com/RWjAdj for more.

Frozen grants: A ¥270-billion (US\$3-billion) funding programme in Japan has been put on hold because of a wholesale budget freeze by the country's new government. The Funding Program for World-Leading Innovative R&D on Science and Technology was created this spring by Japan's former ruling party, the Liberal Democratic Party, as part of a supplementary budget. Thirty research groups were quickly chosen to share the money ahead of the 16 September transfer of power to the Democratic Party of Japan, but none is yet assured of the promised funds.

Chemical regulation: The US Environmental Protection Agency (EPA) laid out White House-backed principles for a radical reform of US legislation regulating toxic chemicals, at present controlled by the 1976 Toxic Substances Control Act. EPA administrator Lisa Jackson said the act had proved to be “an inadequate tool” for protecting the public. She wants to strengthen the EPA's authority to clamp down on dangerous chemicals, and for chemical manufacturers routinely to give the agency toxicity data. The American Chemistry Council, which represents US chemical



ISLANDS SUFFER DUAL SHOCKS

Major earthquakes in the Pacific and Indian oceans struck Samoa and Indonesia last week. An earthquake measured by the US Geological Survey as a magnitude 8.0 triggered a tsunami off Tonga and Samoa that killed at least 190 people and left thousands homeless. The death toll for the magnitude-7.6 earthquake that hit just hours later off the Indonesian island of Sumatra may reach thousands. Experts said that the two earthquakes were not related to each other. See go.nature.com/qkxBhD for more.

manufacturers, says it welcomes the reform. A congressional bill is expected soon.

Biosecurity: The US government should grade microorganisms and toxins according to their risk as potential “biothreat” agents, and regulate them accordingly. That was the recommendation of a National Research Council report released last week, entitled *Responsible Research with Biological Select Agents and Toxins*. Currently, research on 82 human, plant and animal pathogens (called select agents) is monitored under a 1996 law that requires the same security procedures for all of them. The report also called for funding of regular, independent evaluation of the programme governing research into select agents.

European research: Control of Europe's research funds should be devolved from the European Commission to agencies with “arms-length” independence, an

advisory board has suggested. In a report published on 6 October, the European Research Area Board also said that research's share of the European Union's budget should triple to 12% by 2030, with half of those funds going towards basic research. In the most recent funding round, research was allocated €50 billion (US\$74 billion) for 2007–13, of which €7.5 billion went to basic research. See go.nature.com/yVLwT3 for more.

NUMBER CRUNCH

12,000%

The potential rise in India's nuclear capacity, from 3.8 gigawatts today to 470 gigawatts by 2050, in expansion plans announced by the country's prime minister, Manmohan Singh.

(The Times)

● FACILITIES

Dam settlement: After a bitter and lengthy controversy over water management, four hydroelectric dams on the Klamath River in Oregon and California will be removed to restore salmon runs. PacifiCorp, the Portland-based utility in Oregon that owns the dams, announced the draft agreement on 30 September after almost a decade of negotiations with federal agencies, farmers, states, conservation groups and Native

American tribes. Removal of the dams will not begin until 2020, and is dependent on government approval as well as full federal and state funding.

● BUSINESS

Firing frenzy: Sequenom, a biotechnology firm in San Diego, California, has cleared out its top executives after an internal investigation found lax oversight of faulty research. The company revealed in April that data supporting a prenatal screen for Down's syndrome were "mishandled" and could not be relied on (see *Nature* 459, 23; 2009). Last week the company fired chief executive Harry Styllis and Elizabeth Dragon, senior vice-president of research and development. Chief financial officer Paul Hawran and another unnamed executive resigned, and three research scientists also had their contracts ended.

● RESEARCH

Research statistics: China has become the second-largest producer of academic research papers in the world behind the United States, according to a study produced annually for the UK government. Leeds-based analysts Evidence Ltd revised its figures to show that China's output overtook the major European Union states in 2006, although on measures of citation share and research strength (citation share across 10 different fields) it remained lower. The achievement had not been noted until now, and follows an increase in the volume of papers

SOUND BITES

"That is one of the things that wakes me up in the middle of the night."

Francis Collins, the newly minted director of the US National Institutes of Health, tells *Nature* how he feels about trying to ensure that the agency won't suffer financially when its stimulus money is spent. See go.nature.com/h15ch6 for the full interview.

might see their wheat production almost halve. See go.nature.com/VYzOWI for more.

Presidential visit: US President Barack Obama visited the National Institutes of Health in Bethesda, Maryland, on 30 September (pictured). Obama toured a National Cancer Institute lab, where he was treated to video images of healthy and cancer-riddled brains, and praised the research made possible by the agency's spending of \$5 billion out of the \$10.4 billion it got in economic stimulus funds. The agency had raced to disburse the money by that same day, the end of the government's fiscal year.



Carbon cuts: Carbon dioxide emissions could fall by 3% worldwide this year because of the global economic crisis, the International Energy Agency predicted in a teaser from its upcoming World Energy Outlook 2009 report. The excerpt was released on 6 October to coincide with United Nations climate talks in Bangkok.

● AWARDS

Nobel winners: Elizabeth Blackburn, Carol Greider and Jack Szostak shared the 2009

THE WEEK AHEAD

9 OCTOBER

NASA's Lunar Crater Remote Observation and Sensing Satellite will crash into a crater near the Moon's south pole, in the hope of disturbing and detecting ice.

► <http://lcrs.arc.nasa.gov>

15-16 OCTOBER

'The ambitions of Europe in space' — European policy-makers, financiers, space scientists and industrial representatives converge on Brussels to discuss the region's space programme.

► www.spaceconference.eu

15-25 OCTOBER

Canada's Perimeter Institute for Theoretical Physics (see *Nature* 461, 462-465; 2009) hosts — and webstreams — the 'Quantum to cosmos' festival in Waterloo, Ontario.

► www.q2cfestival.com

Nobel Prize in Physiology or Medicine, for their discoveries of how chromosomes are protected by telomeres and the enzyme telomerase. The physics prize went to Charles Kao, for his work on how light can be transmitted through optical glass fibres; and to Willard Boyle and George Smith for their invention of the charge-coupled device (CCD) sensor. The chemistry prize was yet to be awarded as *Nature* went to press. See page 706 and www.nature.com/news for more.

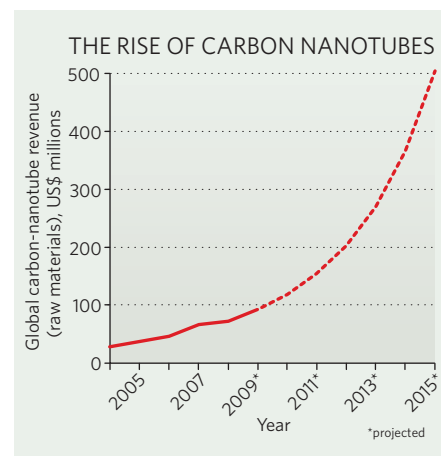
BUSINESS WATCH

In September, French company Arkema became the latest carbon-nanotube manufacturer this year to announce plans for a drastic scaling up of production. Despite the materials' present reliance on the mixed fortunes of the automobile industry, the market for carbon nanotubes as raw materials looks set to grow rapidly. Revenues could reach US\$500 million by 2015 (see chart), predicts Jurrón Bradley of analysts Lux Research.

"In 2009 we have seen a lot of players announce major expansions," Bradley says. Market leader CNano, based in Santa Clara, California, announced in June that its Beijing facility is now producing 500 tonnes of multiwalled nanotubes annually; Arkema's plant in Mont, France,

should be turning out 400 tonnes per year by 2011. If companies such as Germany's Bayer MaterialScience, headquartered in Leverkusen, and Tokyo-based Showa Denko follow through with similar plans, global nanotube production will have doubled by 2011 from around 800 tonnes per year at present.

Many of these raw nanotubes are multiwalled, and are used to make light, strong, composite materials. But a host of smaller companies — such as Nanocyl, based in Sambreville, Belgium — process them into intermediate products for antistatic coatings, for example, sensors for gas detection and electrode material for batteries — and, on the horizon, for touch-screen displays.



Fossil rewrites early human evolution

Ethiopian find dates back 4.4 million years.

A 17-year investigation into a fossilized early human skeleton from Ethiopia culminated last week with 11 papers published in *Science*.

Detailed descriptions of the skeleton, of a fairly complete 4.4-million-year-old female, show that humans did not evolve from ancient knuckle-walking chimpanzees, as has long been believed. The new fossils of *Ardipithecus ramidus* — known as 'Ardi' — offer the first substantial view of the biology of a species close to the time of the last common ancestor shared by humans and apes. Like modern humans, Ardi could walk upright (see depiction, right) and didn't use her arms for walking, as chimps do. Still, she retains a primitive big toe that could grasp a tree like an ape¹.

Previously, the oldest near-complete skeleton of an early human was the 3.2-million-year-old *Australopithecus afarensis* skeleton known as Lucy, also from Ethiopia. Because Lucy had many traits in common with modern humans, she didn't provide much of a picture of the earlier lineage between apes and humans, says Alan Walker, a biological anthropologist at Pennsylvania State University in University Park. The new *A. ramidus* "is so much more important — and strange", he says.

The earliest *Ardipithecus*, *A. kadabba*, lived around 5.8 million years ago in Ethiopia². The other oldest known hominids are *Orrorin tugenensis*, from about 6 million years ago in Kenya³, and *Sahelanthropus tchadensis*, from at least 6 million years ago in Chad⁴ (see graphic).

In addition to describing the fossils, the *Science* papers provide details about the geol-

ogy and palaeoenvironment of the discovery site, in the Afar desert 230 kilometres northeast of Addis Ababa. The research team, known as the Middle Awash Project, involves 70 investigators, 47 of whom are authors on the papers.

In 1992, team member Gen Suwa found the first specimen of *A. ramidus* near the Ethiopian village of Aramis. Within two years, enough fossils had been found to produce the first article that named and sketchily described the animal, from a total of 17 fossils⁵.

Some researchers have complained how long it has taken to publish work about the fossils. But Berhane Asfaw, a co-director of the Middle Awash Project at the Rift Valley Research Service in Addis Ababa, says: "We weren't interested in how many papers we could publish. Our interest was in the full chain of information; that produces the power of the work."

From more than 135,000 vertebrate bone or tooth pieces, the team identified 110 as being from *A. ramidus*, representing a minimum of 36 individuals. The fossils come from a sediment layer sandwiched between two layers of volcanic rock known as tuff — each dated to 4.4 million years ago, says a team led by Giday WoldeGabriel, of Los Alamos National Laboratory in New Mexico. Fossils in the sediments include plants, pollen, invertebrates and birds, which helped to pinpoint the woodland environment where Ardi lived.



Years of field work uncovered Ardi's skull, teeth, arms, hands, pelvis, legs and feet — all of which had to be painstakingly prepared. Ardi's skull was recovered crushed in more than 60 pieces that were broken and scattered about. The bone was poorly fossilized — so soft that each piece had to be moulded in a silicon rubber cast then digitized by computed tomography scans.

Ardi's hands and wrists don't show several distinctive chimpanzee characteristics, such as some larger bones and a tendon 'shock absorber' system to withstand bodyweight, says team member Owen Lovejoy of Kent State University in Ohio. The foot,

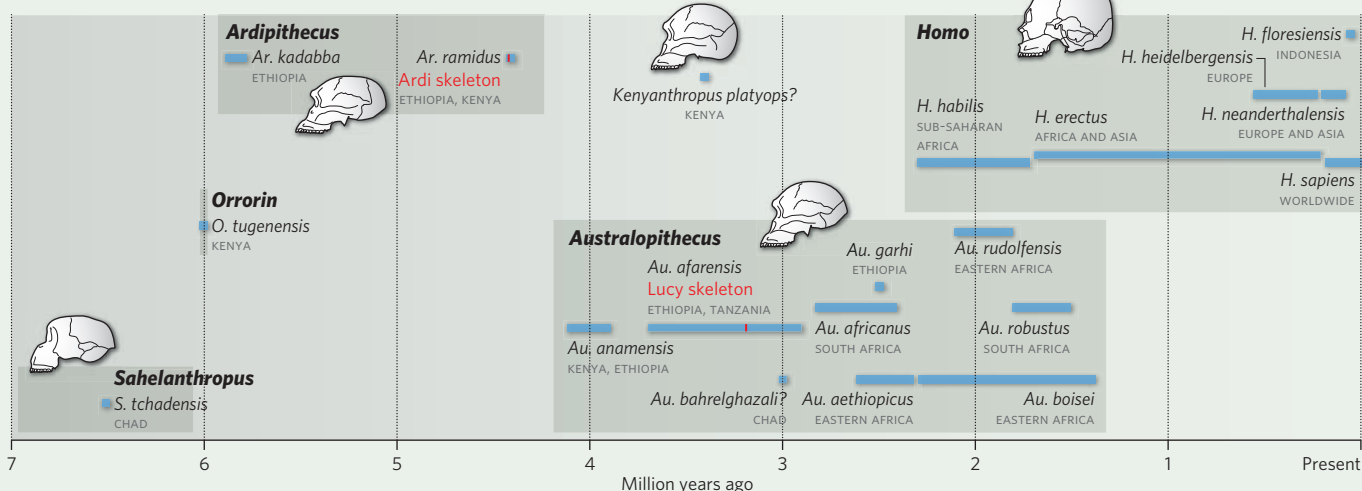
with its big toe sticking out sideways, would have allowed Ardi to clamber in trees, walking along branches on her palms. And her teeth show no tusk-like upper canines, which most apes have for weapons or display during conflict. "This is a major feature showing that Ardi is not in the lineage of modern chimps," Suwa says.

Rex Dalton

1. White, T. D. *et al. Science* **326**, 75–86 (2009).
2. Haile-Selassie, Y. *Nature* **412**, 178–181 (2001).
3. Senut, B. *et al. C. R. Acad. Sci. Paris Ser. IIa* **332**, 137–144 (2001).
4. Brunet, M. *et al. Nature* **418**, 145–151 (2002).
5. White, T. D., Suwa, G. & Asfaw, B. *Nature* **371**, 306–312 (1994).

For a longer version of this story, see go.nature.com/gSAuY5

KNOWN HUMAN ANCESTORS

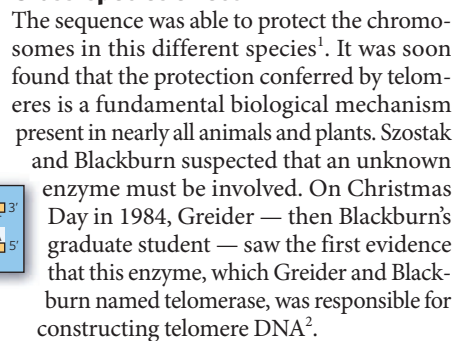




FEET TO RIGHT: M. BROBST/AP PHOTO; M. CROLL /AP PHOTO; M. WILLSON

Prize for physiology or medicine awarded for uncovering role of telomeres.

Their research revealed a fundamental aspect of how DNA, packed into chromosomes, is copied in its entirety by the DNA polymerase enzyme during cell division. The ends of the chromosomes are capped by telomeres, long thought to have a protective function (see ‘Chromosome caps’). Without them, the chromosomes would be shortened during each cell division, because DNA





FUTURE OF HIV VACCINE UNCLEAR

Puzzles to solve before trials can move forward.

www.nature.com/news

PUNCHSTOCK

They worked out that telomerase provides a platform enabling DNA polymerases to copy the entire length of the chromosome without missing the ends. Greider and Blackburn also showed that telomerase contains a key RNA sequence that acts as a template for the telomere DNA³, which attracts proteins to form a protective cap around the ends of the DNA strands.

Telomeres themselves shorten with repeated cell division, making up a key part of the cell's ageing mechanism. Low telomerase activity and telomere shortening speed up ageing, whereas incessantly dividing cancer cells often have high telomerase activity and maintain their telomere length. Cancer therapies directed against telomerase are now being tested in clinical trials.

But there is still a lot of basic biology to discover — such as how telomerase activity is regulated at individual telomeres, and how telomeres manage to avoid the attentions of DNA repair enzymes which seek out breaks in DNA and restitch the torn ends.

Blackburn and Greider become only the ninth and tenth female scientists to win the physiology or medicine prize since it was first awarded in 1901, and it is the first time that two women have been recognized in a single prize. Indeed, telomere research is unusually dominated by women. "It is hard to find a male among us," says David Shore, a cell biologist at the University of Geneva, Switzerland. "And two main reasons are Liz and Carol — they created the field and have been role models."

Blackburn has also been involved in the politics of science, serving on the US President's Council on Bioethics from 2002 until she was dropped in 2004 after criticizing the restrictions on human embryonic stem-cell research imposed by then President George W. Bush.

Lea Harrington, Greider's first graduate student, who is now at the Wellcome Trust Centre for Cell Biology at the University of Edinburgh, UK, says that her four years in Greider's laboratory at Cold Spring Harbor, New York, were "electric. We all realized what an exciting time it was — so many questions being answered about the composition of telomerase, how it worked and its relevance to human biology."

Alison Abbott

1. Szostak, J. W. & Blackburn, E. H. *Cell* **29**, 245–255 (1982).
2. Greider, C. W. & Blackburn, E. H. *Cell* **43**, 405–413 (1985).
3. Greider, C. W. & Blackburn, E. H. *Nature* **337**, 331–337 (1989).

Nobel Prize in Physics awarded to light pioneers

Two technologies that revolutionized science, computing and communication have secured their developers a share of the Nobel Prize in Physics.

Charles Kao of the Chinese University of Hong Kong has won half the prize for his role in developing fibre-optic cables. The other half is shared by Willard Boyle and George Smith of Bell Laboratories in Murray Hill, New Jersey, for their development of the charge-coupled device (CCD), an electronic chip that converts light into a digital signal.

In 1969, Boyle and Smith developed a chip that could transform light into an electronic signal. The duo used newly discovered metal oxide semiconductors that could convert photons into a flow of electrons, which could be read from the edges of the chip and used to recreate the image. The ability to digitally capture light has found application in nearly every field of science — particularly astronomy. "Basically, they revolutionized optical astronomy," says Mark Casali, head of instrumentation at the European Southern Observatory in Garching, Germany. Before the advent of CCDs, astronomers were imaging stars using photographic plates, which were less sensitive and less precise than their digital successors, Casali says. Using CCD cameras, astronomers have been able to discover faint galaxies and even see fluctuations in a star's

light created by an orbiting planet.

The detectors also made space-based astronomy a reality, says Matt Mountain, director of the Space Telescope Science Institute in Baltimore, Maryland, which coordinates science for the Hubble Space Telescope. "It made telescopes like the Hubble possible," he says. "You could now put large electronic detectors in space that could beam down digital pictures of some of the faintest objects human beings have ever seen."

Fibre optics has had an equally impressive impact on science, not least by facilitating collaboration on a global scale. But the transmission of data over thousands of kilometres seemed a distant dream when Kao first began his work on fibre-optic cables. Back then, fibres could carry light only a few metres by total internal reflection before the signal

faded. Kao and his colleagues at Standard Telecommunication Laboratories in Harlow, UK, worked out that impurities, mainly iron ions, were causing the loss. Kao identified an alternative material — fused silica — that could carry light over much greater distances without significant loss. The work ultimately led to the billion-kilometre-long network of fibre-optic cables that span the globe today.

Fibre optics will also have a pivotal role in the world's largest science experiment, the Large Hadron Collider (LHC) at CERN, Europe's particle-physics centre near Geneva, Switzerland. The LHC's largest detectors create around a million gigabytes of raw data every second. The cables then shepherd the data to nearby servers and on to thousands of scientists in 33 countries through an ultrafast computer grid. "The whole infrastructure is based on optical fibre," says Ian Bird, the grid's project leader. "There's no way that our data rates could be sustained without it."

Geoff Brumfiel



Charles Kao: fibre-optic cables.

REUTERS



Willard Boyle and George Smith invented charge-coupled devices.

ALCATEL-LUCENT/BELL LABS

X-ray free-electron lasers fire up

California's project has the lead as its facility goes live, but Europe aims for its own rapid-fire device.

HAMBURG

Heinz Graafsma is tired of the “pretty, but useless” images of proteins that regularly adorn the pages of journals such as *Nature*. “Chemistry depends on changes,” says Graafsma, the head of detectors for photon science at DESY, Germany’s high-energy physics laboratory in Hamburg. “The static world is boring.”

Get ready for the movies. A new generation of light sources — including one just completed in California, one under construction in Japan and one being built outside Graafsma’s office — are getting set not only to put atoms and molecules under the spotlight, but also to illuminate their dynamics.

The devices, called X-ray free-electron lasers, produce flashes of X-ray light with angstrom-level wavelengths — small and coherent enough to image individual atoms. The flashes are also more intense than any created before — stuffed with enough photons to create and study extreme states of matter such as plasma.

But perhaps most importantly, the bursts of light are short — just hundreds of femtoseconds long, the time it takes for light to cross a human hair. Pulses as brief as this can record functions, not just forms: the folding of a protein, the action of a catalyst, the splitting of a chemical bond.

“That is the revolutionary thing,” says Joachim Stöhr, director of the Linac Coherent Light Source (LCLS) at the SLAC National Accelerator Laboratory in Menlo Park, California. The US\$420-million machine, the first free-electron laser in the world to operate at wavelengths this short, began its first experiments last week.

The new devices will outgun the workhorse of the past half-century: the synchrotron, in which beams of electrons, whipped around in a circle, emit bursts of X-ray radiation. Interest in synchrotrons is still high; the number of users at the four major US synchrotron facilities rose



In California, SLAC researchers calibrate the magnets at the Linac Coherent Light Source.

from 6,009 to 8,492 between 2000 and 2008. But these facilities are starting to reach fundamental limits. Some experiments require many photon ‘hits’, and these can require weeks, if not months, at even the brightest synchrotrons. In addition, synchrotron pulses are limited to the picosecond regime, a thousand times longer than free-electron laser bursts. Like using a camera with a slow shutter speed, images of always-jittery molecules end up fuzzy.

Just as synchrotron rings were first built for particle smashing, free-electron lasers also depend on a tool borrowed from particle physics: the linear accelerator or ‘linac’. The LCLS uses the 43-year-old SLAC linac, in which bunches of electrons are accelerated through a 1-kilometre-long tunnel along a path so tightly focused that Earth’s curvature and weak

magnetic field have to be taken into account.

The electron bunches then reach a 130-metre-long section of undulators — magnets that ‘wobble’ the electrons and coax them into emitting X-rays. The wiggles are tuned to the wavelength of the emitted light, creating a feedback mechanism: electromagnetic fields from the X-rays act on the electrons, concentrating them into small, tight groups that emit amplified light in unison.

On 10 April, LCLS engineers successfully tested this crucial idea, which was first proposed in 1971 (J. M. J. Madey *J. Appl. Phys.* **42**, 1906–1913; 1971). LCLS project director John Galayda remembers seeing a sudden surge in light on that day as the electron beam passed the tenth undulator and the amplification began to occur.

Now it’s time to start using the beam. After a summer spent commissioning the first instrument, the first team arrived at 7:30 a.m. on 1 October for five straight days of data collecting. A gas jet shot atoms of neon into the oncoming beam pulses so that the scientists could study what happens when electrons from the atom’s innermost shell are stripped away. “It’s like peeling an onion from the inside out,” says instrument scientist John Bozek.

One of the most anticipated applications for these tiny spotlights (see ‘Science by the

Science by the femtosecond

Atomic physics: exploring how X-rays rip electrons from the inner shells of atoms.

Warm dense matter: creating and studying states of matter that lie between solids and plasmas, found in the interiors of planets and cool stars.

Single particles and biomolecules: eliminating the need for crystallography, which is the main bottleneck in describing complex biological structures.

Femtochemistry: making movies of chemical bonds

being made and broken, of crystals melting and of nanometre-scale droplets nucleating.

Nanometre-scale dynamics in condensed matter:

probing how proteins fold and polymers twist. **E.H.**

femtosecond') will be imaging single biomolecules. At synchrotrons, proteins have to be crystallized so that the lattice-like structure of many identical proteins clarifies the image made by the relatively incoherent X-rays. But some targets, such as viruses and the proteins found in cell membranes, are notoriously difficult to crystallize.

Expanding the number of described protein structures will be important. But DESY director Helmut Dosch says that more surprises will come from descriptions of how those structures move. Many disorders, such as Alzheimer's disease, arise when there is a problem in the way in which proteins fold. "You have to understand what drives the folding," he says.

There is a drawback to free-electron lasers: there are few chances to work with a piece of the spotlight. At circular synchrotrons, light can be siphoned off to experiment stations at regular intervals — the new PETRA III ring at DESY, for example, has 14 stations that can hold up to 30 instruments. But the straight-shot LCLS has just one station holding a single instrument, and work must proceed in series. "The available time is small and the amount of exciting science is large," says Jerome Hastings, head of the LCLS science department. Eventually, SLAC plans to install switching stations, so that light pulses can be diverted to each of six planned experiments.

That's one reason why scientists at DESY think there will still be plenty of work left to do when the European X-Ray Free-Electron Laser (XFEL) is completed in 2014. (Japan also hopes to complete a free-electron laser in 2010, next to its SPring-8 synchrotron in Harima.) An agreement, expected by the end of this year, will formalize the 13-nation, €1.1-billion (US\$1.6 billion) XFEL project.

The XFEL will navigate the terrain beneath

urban Hamburg. Germany is footing up to 60% of the construction costs and is pressing ahead with construction, which began in January.

In August, men in orange coveralls stacked reinforcing bars and electric cabling in a 40-metre-deep chasm, having scooped away hundreds of thousands of cubic metres of dirt and the occasional Second World War mortar shell. From this pit at the edge of DESY, tunnels bearing the linear accelerator will burrow northwest to the town of Schenefeld, 3.4 kilometres away.

The tunnelling is the most expensive component of the XFEL's construction, and much of the 5-year lead that the LCLS has over the XFEL can be attributed to the recycling of SLAC's existing linear accelerator. Galayda says that the Californian project would have cost at least \$300 million more if the team had had to dig a tunnel and build an accelerator from scratch.

But the XFEL has its own trump card: its accelerator will use cryogenically cooled superconducting cavities, allowing the electron bunches to be fired off more quickly. Whereas the LCLS is limited to 120 bursts of light per second, the XFEL will release 30,000.

Massimo Altarelli, XFEL's designate director, says the machine's superior repetition rate will be particularly important in experiments involving dilute targets — a biomolecule floating in a solvent, say — where the chance of registering a photon 'hit' may be slim. But LCLS scientists contend that, most of the time, all of the extra firepower will be wasted: the light will be discarded, absorbed by lead walls.

Stöhr says he is glad that the LCLS is up and running first, but adds that there will be plenty of important science for the XFEL. "There's more than one winner," he says. ■

Eric Hand

"The available time is small and the amount of exciting science is large."

E. HAND

In Germany, work is well under way on the European 'XFEL' light source.





Choren Industries has opened a plant in Germany to produce its synthetic biofuel SunDiesel.

From plant to power

The last of four weekly articles looks at making liquid fuels direct from biomass.

THE BUSINESS OF BIOFUELS

Petrol might yet survive the green revolution. Some investors are taking seriously the concept of 'green gasoline' — transforming the woody remains of plants into exact replicas of today's transportation fuels.

Many see promise because, unlike other biofuels, this product would blend smoothly into today's petrol-driven infrastructure. "This is one I like. It's got a chance of making it," says Lanny Schmidt, a chemical engineer who works on combustion processes and alternative fuels at the University of Minnesota in Minneapolis.

Yet this 'biomass-to-liquid' approach is one of the least known in the biofuels portfolio, and barely makes a dent in alternative fuel quotas. A report by the US Foreign Agricultural Service estimates that in 2009 biomass-to-liquid fuels will make up just 2,000 tonnes of oil equivalent for road transportation in the European Union. The figure for bioethanol is 2 million tonnes, and for conventional fossil fuels it is more than 310 million tonnes. The report concludes that the technology for biomass-to-liquid fuels is "in its infancy and will take some years before it reaches a significant volume".

At least one major oil company has dabbled in the field. In 2008, Royal Dutch Shell invested an undisclosed amount in Virent Energy Systems, a company based in Madison, Wisconsin. The collaboration aims to improve Virent's processes to take sugars generated from cellulosic waste and catalytically react them with water to produce fuel molecules. "It's a premium high-octane [petrol] we're generating," says Randy Cortright, the company's founder and chief

technical officer, who co-invented the technology with chemical engineer James Dumesic of the University of Wisconsin-Madison.

But Virent currently produces about a litre a day, and getting that to more significant amounts will take time. "By the end of this year we will have a larger-scale pilot plant capable of 40,000 litres a year," says Cortright. A full-scale plant is some three to six years away.

The technologies required are known but need refinement: they rely on breaking down biomass, such as sugar molecules, which can then be handled in conventional refineries to produce petrol, diesel or jet fuel. But the catalysts needed to convert the biomass to useful hydrocarbons are still being developed, as are ways to break down the biomass so it can be processed.

George Huber, a chemical engineer from the University of Massachusetts in Amherst, heats biomass so that it decomposes and releases volatile sugars, which are then passed over a zeolite catalyst to form the aromatic molecules benzene, xylene and toluene. These molecules are also extracted from crude oil, and xylene and toluene can be blended with other substances to make petrol. A mixture of these molecules can be produced for less than US\$0.46 per litre, says Huber, who has started a company called Anellotech, to develop the technology further.

Still, investors need patience, he says: "You make your money on volume." To go from lab-based processes, such as his, to the pilot stage takes three or four years; to scale up to a large

demonstration plant will take another three or four years and hundreds of millions of dollars, he estimates.

Nevertheless, some money is flowing into the field. Virent has raised \$30 million in venture financing and has commitments of \$40 million from industry and government funding. Amyris Biotechnologies, a firm in Emeryville, California, that engineers microbes to increase biofuel yield, has amassed \$140 million since 2006 from high-powered investors including Khosla Ventures and Kleiner Perkins Caufield and Byers.

And late last month LS9, based in South San Francisco, attracted \$25 million from oil and gas company Chevron. LS9 uses specially designed microbes to chew up biomass to produce hydrocarbons that can be refined as usual.

"The promise of these fuels is that oil companies will be able to use them very easily," says Harry Boyle, an analyst at London-based New Energy Finance. "The exit, if you're a venture-capital player, is huge and very exciting."

But it's a long way to a profitable exit. For instance, biofuels producer Dynamotive Energy Systems, based in British Columbia, Canada, posted net losses of Can\$1.5 million (US\$1.4 million) in the second quarter of 2009.

Some are looking to federal loans to help out. Terrabon, based in Houston, Texas, announced in July that it had made organic salts from biomass, then turned them into petrol with Valero Energy Corporation, a refining company headquartered in San Antonio, Texas.

"The promise of these fuels is that oil companies will be able to use them very easily."

C. KOALL/GETTY IMAGES



Q&A: A CONSERVATIVE PLAN FOR UK SCIENCE

Adam Afriyie on research under a potential centre-right government.

www.nature.com/news

Terrabon, which uses technology developed at Texas A&M University in College Station, Texas, plans to build a bigger plant in Port Arthur, Texas, that can process 55 tonnes of biomass a day, producing 4.9 million litres of fuel a year. It has applied for a \$25-million grant from the US Department of Energy to build this plant, but if it doesn't get the grant it will invest even more itself and make the plant even larger — to process up to 220 tonnes of biomass each day. If paying for the whole plant, says chief financial officer Malcolm McNeill, “you might as well go to the real size”.

Meanwhile, the processing company UOP, based in Des Plaines, Illinois, has developed a pyrolysis technique that heats biomass to release oil. More work is needed to develop that oil into a fuel, but the technology is already being licensed by UOP's joint venture with Canadian company Ensyn Technologies in Ottawa, Ontario. “What this technology has lacked is the economic drivers to make it happen,” says Graham Ellis, UOP's business manager for renewable energy and chemicals. UOP wants to help existing refineries to license its upgrading technology to use in existing infrastructure.



At Virent, researchers are engineering microbes to increase biofuel yield.

In Germany, the car-maker Volkswagen, based in Wolfsburg, is leading a €13.6-million (US\$20-million) project intended to eventually produce 200,000 tonnes per year of liquid fuels from biomass. The processing will be done by Choren Industries in Freiberg. Choren has separately amassed investments of €140 million, mainly from individual investors, although minority shareholders include Shell Deutschland Oil, Daimler and Volkswagen. It

is now commissioning a new plant in Freiberg that will have a nominal capacity of 18 million litres of synthetic biofuel per year, all of which will be sold to Shell.

Some producers think they can eventually become competitive by offering a lower-cost product than many other first-generation biofuels. Raw-material costs for synthetic biofuels, says Choren spokeswoman Ines Bilas, can be around 40% of total costs, compared with nearly 90% for biodiesel made from rapeseed oil.

The fuel's adaptability may also help it to catch up with other, more established biofuel alternatives. “You really can make [petrol] from sorghum or municipal waste,” says McNeill.

But for now, its future rests with process engineers and how well they can streamline the path from woody waste to liquid fuel. ■

Katharine Sanderson

Correction

The News story ‘Climate burden of refrigerants rockets’ (see *Nature* **459**, 1040–1041; 2009) cited an incorrect year for when hydrofluorocarbon emissions were predicted to reach between 5.5 billion and 8.8 billion tonnes of carbon dioxide equivalent annually. The year is 2050, not 2010.



HIT OR MISS?

Genome-wide association studies have identified hundreds of genetic clues to disease.

Kelly Rae Chi looks at three to see just how on-target the approach seems to be.

Five years ago human geneticists rallied around an emerging concept. Technology had granted the ability to compare the genomes of individuals by looking at tens of thousands of known single-letter differences scattered across them. These differences, called single nucleotide polymorphisms or SNPs, served as reference points or signposts of common variation between individuals. The idea was that common variants in the genome might contribute to the genetics of common diseases.

Genome-wide association studies (GWAS) could scan SNPs in thousands of people, with and without a disease. When a DNA variant can be associated with the risk of developing a disease, it signals that something in that area of the genome might be partly responsible. With such 'hits' would presumably flow a better mechanistic understanding of disease, genetic-testing abilities and even treatment.

"Many researchers really grabbed on to the common variant hypothesis, and in some cases it worked," says Jonathan Haines, director of the Vanderbilt University Medical Center's Center for Human Genetics Research in Nashville, Tennessee. But, he adds, "it hasn't panned out to be as pervasive an explanation as we thought".

Here are the stories of three hits. One provides a near perfect example of the positive outcome that this sort of unbiased approach can have. One reveals that without biological context the findings can be hard to interpret. And the third demonstrates that GWAS in their current form can't cope well with some common traits.

A direct hit in haemoglobin

In 2007, researchers reported on genome-wide scans of healthy adults looking for SNPs associated with very high or

very low levels of fetal haemoglobin. Among several hits were variants of a gene on chromosome 2 called *BCL11A* (ref. 1). This finding, quickly replicated in multiple populations, generated a lot of excitement.

Fetal haemoglobin is a remnant of embryonic development. For most people, the fetal version of this crucial oxygen-carrying protein drops off after birth as the adult version kicks in. Some people retain relatively high expression, which seems to have no effect in healthy adults. But for patients with blood disorders such as sickle-cell disease and β -thalassaemia, those expressing high fetal-haemoglobin levels can be protected from some of the nastier ravages of the disease, such as leg ulcers, severe pain and even death.

GWAS findings often just provide the signpost, a rough coordinate for a causal gene. The SNP signal is often outside gene sequences. The variants in *BCL11A* were a direct hit in a gene, however, and a surprising gene at that. The protein it codes for, which controls the expression of other genes, had been associated with cancer progression, but never with haemoglobin production. A mouse model had even been made in which the gene had been knocked out, but until the GWAS no one had looked at its regulation of blood. "Nobody would have ever dreamed that a gene like this would have any regulatory role in fetal haemoglobin," says Martin Steinberg, a haematologist at the Boston University School of Medicine in Massachusetts. Last year, he and his colleagues replicated the *BCL11A* finding in three different populations with haemoglobin disorders².

Of course, there was functional work to be done. Stuart Orkin's lab at Harvard Medical School in Boston reduced expression of *BCL11A* in cultured blood progenitor cells from humans. Fetal haemoglobin expression went up,

ILLUSTRATIONS BY J. TAYLOR

suggesting that the gene normally acts as a repressor³. In a follow-up study⁴, the group showed that the gene controls the silencing of fetal haemoglobin during development in mice.

But the exact switch mechanism has not been solved, says Orkin. His group has gone on to look for proteins and other genes that the product of *BCL11A* binds to or influences. Results from these studies will inform research that looks for molecules that would interfere with the gene's expression or function and thus serve as potential therapies to activate synthesis of fetal haemoglobin in people with blood disorders.

Steinberg, for his part, hopes to use these and other GWAS findings to refine a computational tool that predicts disease severity or death in people with sickle-cell disease. The hope is to intervene earlier and with more specific treatments.

The verdict: even those who have been generally pessimistic about the outcomes of GWAS consider the *BCL11A* find a win for science. "It's a tour de force illustration of the value of GWAS," says David Goldstein, a geneticist at the Duke Institute for Genome Sciences & Policy in Durham, North Carolina. "You learn something new, you understand the mechanism, and it's biologically and clinically important." As winners go, however, Goldstein says it is on a short list.

In some ways, says Steinberg, the GWAS provided a lucky hit. The researchers built on years of evidence that fetal haemoglobin has a powerful effect on the severity of sickle-cell disease and β -thalassaemia. It was the clear physiological signal — quantity of fetal haemoglobin — that helped researchers to design the GWAS. Other genes and pathways will be found to affect the severity of the disorders, but probably none with the same force as fetal haemoglobin. "We're not going to find another fetal haemoglobin," Steinberg says.

Scoping schizophrenia

Schizophrenia genetics has been a mire of false starts. Scores of candidate gene association studies had identified promising targets, but few held up to further scrutiny. So the excitement around approaching the disease in an unbiased genome-wide study was high. But the first four schizophrenia GWAS reported no statistically significant associations. Then, in research published last year, researchers performed scans in roughly 500 people with the disorder and 3,000 healthy controls. When 12 of the hits that turned up were examined in 16,000 more individuals, a signal started to emerge. Three variants were significant, but only one of them was in a gene, *ZNF804A*, which encodes a protein with unknown function⁵.

Having a potential candidate gave researchers something concrete to work with. One group took

115 healthy people, a little less than half of whom had two copies of the high-risk *ZNF804A* variant, and compared their brain activity using functional magnetic resonance imaging (fMRI), a method that reveals local blood oxygenation and presumably electrical activity in the brain in real time. Those with the variant, they found, had abnormal connectivity between certain brain areas, impairing "the degree in which they talk to one another", says Andreas Meyer-Lindenberg, the director of the Central Institute of Mental Health in Mannheim, Germany, who led the study⁶. Healthy adults with the variant were showing schizophrenia-like brain activity even though they showed no outward signs of disease.

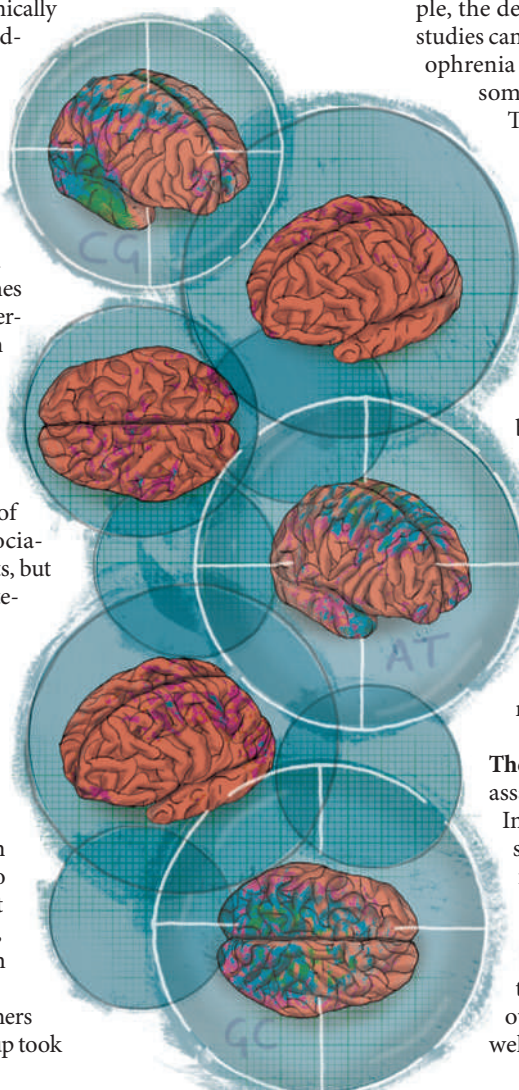
Combining genetic and brain-imaging data to study psychiatric disease is not new. Since 2001, researchers have used the strategy to link imaging data to candidate gene findings in schizophrenia, depression and autism. Meyer-Lindenberg's study is the first to use a genetic loci identified through GWAS for follow-up with fMRI. "We've now applied [the technique] to a variant that has definitive support as being a schizophrenia risk gene. That wasn't available before," says Meyer-Lindenberg.

Part of the problem when seeking schizophrenia-related genes is that, unlike fetal haemoglobin levels for example, the definition of the trait both within and between studies can differ. Also the spectrum and severity of schizophrenia symptoms varies between individuals and are sometimes subjective from a clinical perspective.

That's why fMRI is attractive. The researchers hope to get closer to quantitative measures of psychiatric disorders. "It makes sense to have a biological level of analysis on which these genetic associations can be studied," says Daniel Weinberger, the director of the genes, cognition and psychosis programme at the National Institute of Mental Health in Bethesda, Maryland, who pioneered the method in the 1990s.

The *ZNF804A* association from GWAS has been replicated in some studies but not others. And there are few clues to the mechanism by which this gene might contribute to brain connectivity. It was the group of Michael O'Donovan, a professor of psychological medicine at Cardiff University, UK, that made the initial discovery using GWAS. The team is now carrying out a series of experiments to determine which DNA sequences and other proteins it binds, and how variants might alter gene expression.

The verdict: some have doubts about the combined assault of GWAS and fMRI on psychiatric illness. Imaging data itself isn't the best quantitative trait, says Goldstein, because one three-dimensional fMRI image can contain more than 50,000 picture elements of data, a single trait can be defined in multiple ways. "There hasn't been a sufficient consistency in how those phenotypes are defined," he says. Weinberger and others contend that the imaging paradigms are well established before they are used in imaging



genetics research. "I think it's very important that the phenotypes are well validated — that they are themselves heritable, and that they're related in some way to the underlying neural circuitry," says Weinberger.

Reviews from others studying schizophrenia are somewhat lukewarm. Kári Stefánsson, chief executive of the Icelandic biopharmaceutical company deCODE Genetics in Reykjavik, says he's not completely convinced. "In schizophrenia, the imaging differences are subtle," he says. Nevertheless, he plans to study differences in brain morphology using imaging and GWAS, in people with and without the disorder.

Given the shortage of standout GWAS hits, should researchers continue to use the candidate-gene approach to form the basis for hypothesis-driven imaging genetics work? "It is still a point of debate," Meyer-Lindenberg says.

Sight set on height

Height has produced clearer hits than schizophrenia, but with a less than satisfying punch. In 2007, by analysing the genomes of nearly 5,000 people, researchers were able to see that a variant in a gene called *HMGA2* explains some of height's variability — about 0.3% (ref. 7). Since then, additional GWAS have revealed more than 40 loci involved in height. Added together, these variants account for 5% of the trait's variation. Even a clear quantitative trait doesn't necessarily provide simple answers.

Genes are thought to contribute to roughly 60–80% of the variation in stature, leaving much of the heritability of height unaccounted for by GWAS findings. This 'missing heritability' has been a thorn in the side of the common-disease-common-variant hypothesis (see page 747). "In the field of height," says Haines, "obviously that hypothesis is not completely correct."

But the news isn't all bad. "Optimistic people like me say we didn't know anything about the genetics of height before 2007," says Guillaume Lettre, a geneticist at the Montreal Heart Institute in Quebec, Canada. "Now we have more than 40 loci."

Researchers may have loci, but they have little idea how these contribute to height. As with other traits, many of the associated SNPs fall within the vast regions between genes or within genes whose function is unknown. And with little funding for understanding height variation and scant biological footholds, the field sees very little follow-up of its GWAS leads.

Lettre is collaborating with others with the hope of tying mystery SNPs to genes through animal models. "Basically what we are doing is taking the genes near these markers and looking at the expression of these genes in tissues that are relevant to height," he says. "There are not so many: bone, cartilage and pituitary gland."

He and others are also trying to coax existing height data into revealing stronger associations by grouping hits based on a single molecular pathway. Hong-Wen Deng, at the

University of Missouri in Kansas City, is planning to analyse pathways involved in either bone health or stature. "Many genes which may have small effects for height may not be detected if you analyse them individually," he says. But jointly their effects may be detected. Others are looking at height at various points with the hope that differences in growth curves will reveal larger genetic associations. Researchers have already examined height and growth rates in about 3,500 people from Northern Finland. Of 48 height-associated variants that they tested, 12 were linked with the rate of growth during infancy or puberty⁸.

The verdict: some of the loci implicate molecular pathways already known to be involved in growth and development. A 1995 study had shown that a gene related to *HMGA2* could influence height: mice lacking the gene were shorter, whereas mice with a truncated version developed gigantism⁹.

The *HMGA2* association has been further confirmed by most, but not all, GWAS.

Predicting how hits outside genes will fare is more difficult, and depends to some extent on how close the hit is to the nearest gene. "If you look at the height loci, they are much more likely to be near a gene that causes abnormal skeletal growth, than a similarly sized random set of loci," says Joel Hirschhorn, a geneticist at the Broad Institute in Cambridge, Massachusetts.

But the nearest gene is a poor marker for what is likely to be causal says Goldstein. "Depending on the genetic model for what is causing the association, it could be nearby or not nearby," he says. In some instances changes in DNA act on genes a million bases away. "It really is remarkable that there are hundreds of reported associations, and the number that you can actually track down to an actual cause of the association is probably countable on one hand."

Researchers point to height as a 'model trait' because it is simple to measure and relatively constant compared with phenotypes such as blood pressure or glucose level. Then again, in GWAS of height, tens of thousands of people have been necessary to see the slightest associations. As a model trait, that could be problematic. "In some ways, it is showing us the future for other traits," says Karen Mohlke, a geneticist at the University of North Carolina at Chapel Hill who was involved in some of the initial height GWAS work. "What it means for many other complex traits is that there will be as many loci found, or more."

Kelly Rae Chi is a freelance writer in Cary, North Carolina.

1. Menzel, S. *et al. Nature Genet.* **39**, 1197–1199 (2007).
2. Sedgewick, A. E. *et al. Blood Cells Mol. Dis.* **41**, 255–258 (2008).
3. Sankaran, V. G. *et al. Science* **322**, 1839–1842 (2008).
4. Sankaran, V. G. *et al. Nature* **460**, 1093–1097 (2009).
5. O'Donovan, M. C. *et al. Nature Genet.* **40**, 1053–1055 (2008).
6. Esslinger, C. *et al. Science* **324**, 605 (2009).
7. Weedon, M. N. *et al. Nature Genet.* **39**, 1245–1250 (2007).
8. Sovio, U. *et al. PLoS Genet.* **5**, e1000409 (2009).
9. Zhou, X., Benson, K. F., Ashar, H. R. & Chada, K. *Nature* **376**, 771–774 (1995).

See Editorial page 697, and the human genetics online special at <http://go.nature.com/VqPUE2>.



THE DISAPPEARING NUTRIENT

Phosphate-based fertilizers have helped spur agricultural gains in the past century, but the world may soon run out of them. **Natasha Gilbert** investigates the potential phosphate crisis.

Ten years ago, Don Mavinic was working on a way to get rid of a pesky precipitate that plugs up the works of waste-water treatment plants. Known as struvite, the solid crud forms in pipes and pumps when bacteria are used to clean up sewerage sludge.

Mavinic, a civil engineer at the University of British Columbia in Vancouver, Canada, realized that struvite was more than just rubbish. A combination of phosphate, magnesium and ammonium, struvite contains many of the essential nutrients that plants need. Mavinic has developed a way to remove the precipitate during the water-treatment process and he is now selling it as a 'green' fertilizer. His technology was first used commercially in 2007 in a treatment plant in Edmonton, Alberta, Canada. It has since been exported to a plant in Portland, Oregon, which began using it this year. A sewage works in Derby, UK, successfully tested the technology in September.

Aside from finding a use for a troublesome by-product, the recycling of struvite could also help solve a much bigger problem: the dwindling supply of phosphate rock. All life forms require

phosphorus in the form of phosphate, which has an essential role in RNA and DNA and in cellular metabolism. Every year, China, the United States, Morocco and other countries mine millions of tonnes of phosphate from the ground (pictured above), the bulk of which is turned into fertilizer for food crops. But such deposits are a finite resource and could disappear within the century.

Experts disagree on how much phosphate is left and how quickly it will be exhausted. But many argue that a shortage is coming and that it will leave the world's future food supply hanging in the balance.

"I am starting to think phosphate rock is becoming a strategic material for many countries. In the future it's going to become more and more valuable," says Steven Van Kauwenbergh of the IFDC, an International Center for Soil Fertility & Agricultural Development based in Muscle Shoals, Alabama. Indeed, as political and social tensions build over the reserves of phosphate rock, the world could move from an oil-based to a phosphate-based economy, say some scientists and industry representatives.

"It is a very curious thing that something so important is so poorly understood and so little talked about in the larger political arena," says Arno Rosemarin, a water-resources specialist at the Stockholm Environment Institute who has researched global phosphate use. Although international leaders have not tended to focus on the potential for phosphate shortages, the issue has been proposed for discussion next month at a United Nations meeting on global food security — an indication that it is starting to attract the attention of the international community.

Just decades left?

In many countries, phosphorus is a limiting plant nutrient in short supply in the soil. So farmers add phosphate-based fertilizers to increase agricultural yields. That has spawned a global phosphate-mining industry with sales totalling in the tens of billions of dollars.

The US Geological Survey (USGS) in Reston, Virginia, estimates that around 62 billion tonnes of phosphate remain in the ground (see graphic). This includes 15 billion tonnes of deposits that are mineable at present and others

Y. ARTHUR-BERTRAND/CORBIS



that are not being exploited. The latter are left in the ground mainly because they contain too many impurities — such as cadmium and other toxic metals — or because they are offshore in difficult-to-reach places.

In 2008, 161 million tonnes of phosphate was mined around the world, according to the latest, as yet unpublished, figures from the US Geological Survey. Stephen Jasinski, phosphate-rock commodities expert at the survey, says that demand for fertilizers is predicted to grow by 2.5–3% per year for the next 5 years. If that rate continues, the world's reserves should last for around 125 years.

That is a relatively optimistic timescale, but it is echoed by the International Fertilizer Industry Association in Paris, whose members include 90% of the world's fertilizer producers. Michel Prud'homme, executive secretary of the association's Production and International Trade Committee, says that the industry anticipates that demand for fertilizers will grow at a "fairly moderate rate", slowing by the middle of the century. That would enable reserves to last for at least another 100 years.

But others predict a faster growth in demand for fertilizers, which would deplete phosphate reserves even quicker. The increased use will be driven in part by the rising global population,

which will require food production to at least double by 2050, according to the Food and Agricultural Organization of the United Nations (FAO).

Rosemarin and others say that nations should not rely on the reserves laden with impurities or located offshore because of the costs — both environmental and economic — of extracting usable phosphate. The remaining accessible reserves of clean phosphate rock would run out in 50 years, if growth stays at 3% per year, says Rosemarin.

But the estimates all suffer from a lack of reliable data. Most of the world's phosphate-mining companies are integrated with fertilizer firms and the mines are either owned by the companies or are under state control, says Prud'homme. As a result, it is difficult to get accurate, independent information on phosphate reserves.

Eric Kueneman, deputy director of the FAO's plant production and protection division says, "the reality is we as a public institution don't really know what the industry knows and nor do they know among themselves. To give a reliable answer to the question, 'will phosphates run out?', we need a crystal ball."

The International Fertilizer Industry Association collects data from its members on their existing reserves and on potential upcoming capacity. But some experts question the accuracy of these data because they are supplied by producers who might be disinclined to provide proprietary information that could harm their commercial positions.

No agreement

There is also a lot of uncertainty over the data supplied by governments, which is the case with China and Morocco, says Dana Cordell, who has just completed her doctoral thesis on the effect of phosphate reserves on food security at the University of Technology Sydney in Australia. For example, when China joined the World Trade Organization in 2001, its reported reserves of phosphate rock instantly jumped from just over 2 billion tonnes to nearly 8 billion tonnes¹.

Cordell and Kueneman call for independent data collection on phosphate rock reserves. "Unlike for energy, water or nitrogen, there is no single international organization responsible for phosphate resources. That is very concerning," Cordell says.

The IFDC hopes to generate more solid data about the extent of the world's phosphate resources and reserves. It will soon launch a project that will query phosphate producers,

academics and other minerals specialists to collect extensive data on how much phosphate there is, how pure it is, what might be available in the future and the useful life of existing mines. Van Kauwenbergh, who is leading the project, expects to publish the first round of data in May next year. If the centre secures more funding, he hopes to continue the research for another 5 years.

The USGS figures on phosphate reserves are the most-quoted publicly available information.

But there are problems with them because the agency gets its information from foreign governments, not directly from producers, and it is not independently verified. "We just don't know how good the USGS data are because they are based on second and third-hand information. The figures

change all the time," says Van Kauwenbergh.

Some people who track the phosphate industry say that there is no cause for concern about phosphate running out. "I don't think this is an immediate crisis, but it is something we should be paying attention to," says Jasinski.

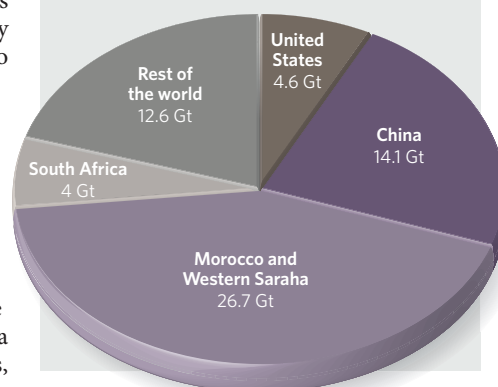
Prud'homme is sanguine about prospects for the future. If demand rises, then so will prices, he says, allowing companies to explore for new reserves and mine those that are harder to reach or from a lower grade of rock. "We feel there are enough reserves to meet food and material needs," he says.

For example, companies have recently begun to investigate deposits in Peru, Australia and off the coast of Namibia that were not previously considered financially viable, says Prud'homme. These resources are not fully taken into account in the most recent USGS figures on world phosphate reserves, he says. And as some existing mines are tapped out, others are opening up in places such as Saudi

"I am starting to think phosphate rock is becoming a strategic material for many countries." — Steven Van Kauwenbergh

THE WORLD'S REMAINING PHOSPHATES

(Gt = gigatonnes)



SOURCE: USGS

Arabia. "I am convinced there are other sources we have not yet found, but it is difficult to say how much impact these will have," he says.

Others are sceptical that further exploration will uncover large new deposits or that they will solve the longer-term problem. "We are not going to find another Morocco," says Jasinski, referring to the country with the biggest remaining reserves.

In the meantime, companies have started to invest in new technologies to exploit the lower-grade and offshore deposits. The impetus for this move into more costly production was the hike in phosphate rock prices in 2008, when the value temporarily spiked at US\$500 per tonne, more than five times the average price in 2007 (ref. 2). Prices had remained comparatively flat for the previous five years. The price hike was due to tight supplies of the rock caused by increased demand for phosphate-based fertilizers in India and China as well as record energy prices. Phosphate prices have since dropped back to their pre-spike levels.

Few alternatives

Despite the investments in unconventional reserves, those deposits may not be viable in the long term. Jan-Olof Drangert, an expert in water and land resources at Linköping University in Sweden, says that lower-grade reserves are "not a solution" if the world wants a sustainable system. Not only will extracting lower-grade phosphates be very expensive, it will also pollute the soils with cadmium, which is highly toxic to plant and animal life even in low doses, he says. "And then there is still the problem of exhausting these lower-grade reserves," he adds.

The increase in demand for fertilizer in 2008 may have been a taste of things to come, especially if demand for food rises as fast as some estimates suggest. The price hike last year "was a huge shock to farmers", says Cordell. Fertilizers had to be rationed in some cases.

"The bottom line is that it will just cost more to eat," says Rosemarin. "There will be no cheap lunches any more."

Making fertilizers go further

No matter how much phosphate is left to be extracted from the ground, cutting down on the use of phosphate-based fertilizers and improving their efficiency could make a significant improvement, says Alan Townsend, a biogeochemist at the University of Colorado in Boulder. "Fertilizer is seen as a cheap insurance policy. Farmers tend to overuse it because they don't want to be caught out," he says.

In the past two decades, the United States and Europe have reduced the widespread

over-application of fertilizers, but that strategy continues to be a problem in other parts of the world, says Townsend. One of the biggest culprits is China⁵, where farmers are applying nearly twice as much fertilizer as is needed in the production of wheat.

Experts disagree, however, on whether excess fertilizer application is actually unwarranted. Tony Vyn, an agronomist at Purdue University in West Lafayette, Indiana, says that the overuse of fertilizers in the European Union and United States has

built up phosphate reserves in the soil. Farmers are now taking advantage of that by applying less phosphate than the crops actually need each year. So the strategy of China's farmers may not be unreasonable, he says.

Other gains toward preserving phosphate resources could come through improved industrial practices. Between 40% and 60% of phosphate is lost when its host rock is converted to fertilizer. Researchers are now looking to reduce that wastage.

N.G.

The uncertainty over the world's phosphate reserves is compounded by the fact that supply is concentrated in just a few hands. China, Morocco, the United States and Russia together hold more than 70% of the global phosphate deposits³, presenting the possibility of "market manipulation", says Amit Roy, president of the IFDC.

Evidence of strategic manoeuvring can already be seen. In March 2004, the United States and Morocco signed a free-trade agreement that covered phosphate rock, among other commodities. In 2008, Morocco exported \$65-million worth of fertilizer to the United States⁴. Although the United States has one of the world's largest phosphate rock reserves, the nation will see a significant drop in production in 25 years when it is estimated that production will peak at its key mines in Florida. The deal with Morocco, says Rosemarin, is aimed at securing the United State's future fertilizer and food supply.

In the case of some finite resources, such as oil, alternatives can be found. But there are currently no substitutes for phosphates. Cutting usage will help to make reserves last longer (see 'Making fertilizers go further').

But most agree that some of the biggest gains will probably be made from the recovery and recycling of phosphates, such as Mavinic's work mining the phosphate deposits inside water-treatment plants. In a back-of-the-envelope calculation, he estimates that if all domestic wastewater facilities in Canada were converted into biological treatment systems using his technology, the country could

produce enough fertilizer to meet about 30% of its current needs.

That pales, however, when compared with a much richer — and more pungent — source of phosphate: the manure generated by dairy and pig farming. Livestock waste contains around five times more phosphate than human waste. And the global livestock population is around 65 billion, more than ten times the human population. There is "enormous potential" for

recovering phosphates from livestock waste, says Mavinic, who has turned his attention to doing just that.

The problem his research team is trying to solve is that phosphates in livestock waste are not in a dissolved form, which is necessary to make struvite. If programmes to

recover phosphates from livestock waste succeed, "the sky is the limit", says Mavinic. "We would probably not have to import any fertilizer into this country."

But all this takes time. Decades may pass before recycling technologies gear up and new supplies of phosphate come on line. At present, nations have expressed little concern over the finite phosphate resource and are eagerly consuming reserves. When solutions do eventually emerge, the world could already be in the grip of a fertilizer and food shortage.

Natasha Gilbert is a reporter based in Nature's London office.

"There is no single international organization responsible for phosphate resources."
— Dana Cordell



Struvite build-up in water-treatment pipes could be a valuable source of phosphate.

1. Rosemarin, A. *Down to Earth* June, 27–31 (2004).
2. USGS Mineral Commodity Summaries: Phosphate Rock (USGS, 2009).
3. IFDC *Global Phosphate Reserves, Resources and Future Production* (IFDC, 2008).
4. <http://www.moroccousafta.com/tradedata.htm>
5. Vitousek, P. M. *et al. Science* **324**, 1519–1520 (2009).

Throw a rock into a quiet pond, and watch the ripples spread out across its surface. This is pretty much what happens when a photon hits the surface of a metal — except that in this case, the ‘ripples’ consist of electrons oscillating en masse and have wavelengths measured in nanometres. Once they are set in motion, these ‘surface plasmons’, as the oscillations are known, can pick up more light and carry it along the metal surface for comparatively vast distances. “A river of light” is how Satoshi Kawata, a physicist at Osaka University in Japan, describes the phenomenon to his students.

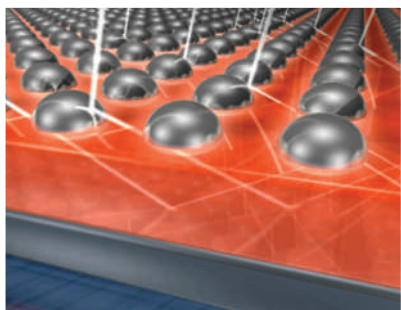
Plasmons can also focus light into the tiniest of spots, direct it along complex circuits or manipulate it many other ways. And they can do all of this at the nanoscale — several orders of magnitude smaller than the light’s own wavelength, and therefore far below the resolution limits of conventional optics.

The result is that plasmonics has become one of the hottest fields in photonics today, with researchers exploring potential applications in solar cells, biochemical sensing, optical computing and even cancer treatments (see ‘Plasmons at work’).

Their efforts, in turn, have benefited greatly from the flowering of nanotechnology in general over the past decade, which brought with it a proliferation of techniques for fabricating structures at the nanoscale — exactly what plasmonics needed to progress from laboratory curiosity to practical applications. “The late 1990s was kind of the turning point” for plasmonics, says Harry Atwater, a physicist at the California Institute of Technology in Pasadena.

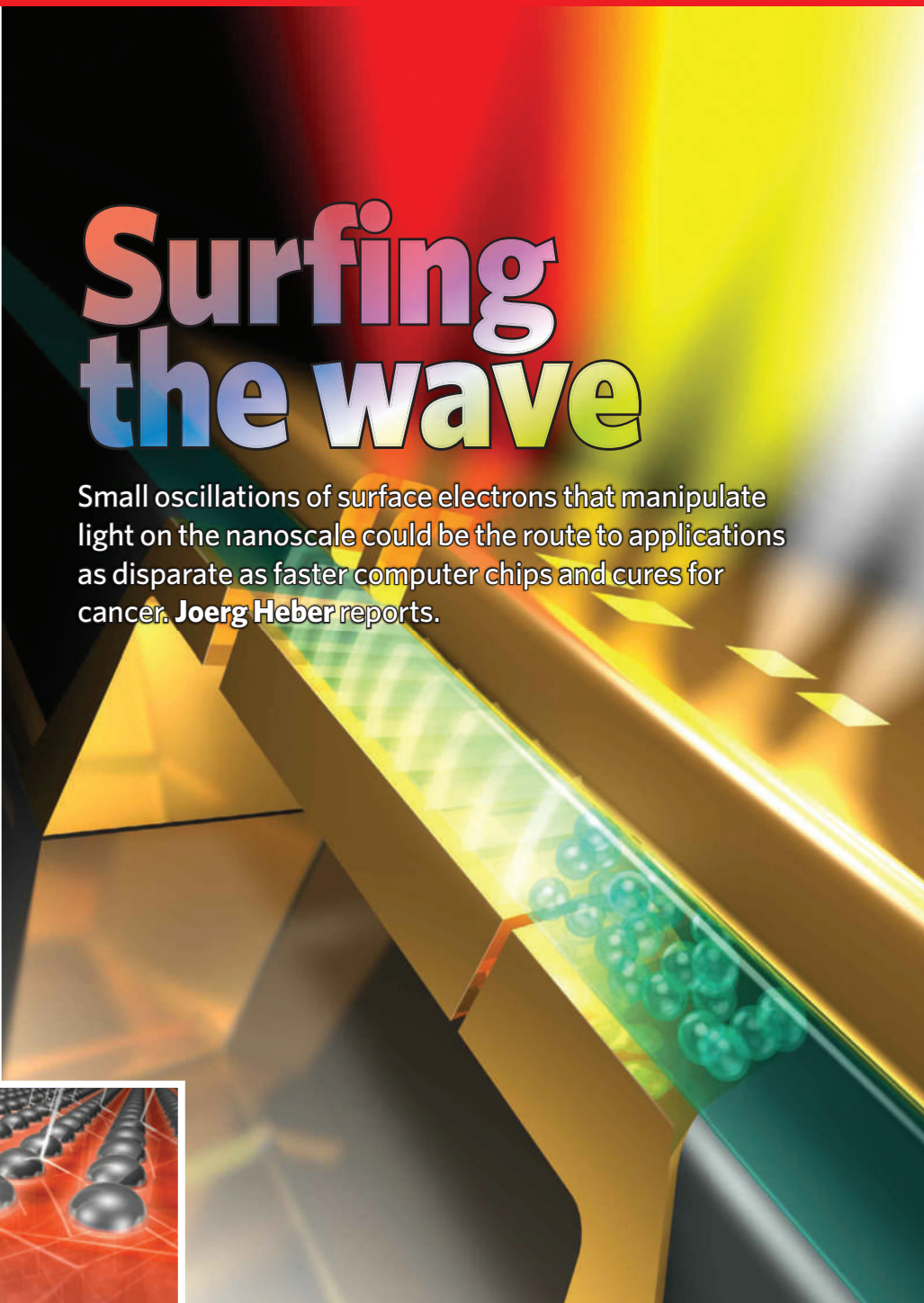
One surprising example of the light-carrying phenomenon was witnessed in 1989 by Norwegian-born physical chemist Thomas Ebbesen, now at the Louis Pasteur University in Strasbourg, France. As he held to the light a thin film of metal containing millions of nanometre-sized holes, he found that it was more transparent than he expected. The holes were much smaller than the wavelength of visible light, which should have made it almost impossible for the light to get through at all. “I first thought, ‘Here was some kind of mistake,’” says Ebbesen.

But it wasn’t a mistake, although it took Ebbesen and his colleagues the better part of a decade to work out what was happening. When the incoming photons struck the metal film, they excited surface plasmons, which picked



Surfing the wave

Small oscillations of surface electrons that manipulate light on the nanoscale could be the route to applications as disparate as faster computer chips and cures for cancer. **Joerg Heber** reports.



Light manipulation: surface plasmons could be generated (above) to help direct light using nanoantennas in devices such as solar cells (left).

up the photons’ electromagnetic energy and carried it through the holes, re-radiating it on the other side and giving the film its transparency¹.

Hole arrays are increasingly finding their way into applications, for example as selective filters for colour sensors. It turns out that the increased transmission through the sheet works only for light around the plasmons’ natural oscillation frequency. But this frequency, which is typically in the visible or near-infrared part of the spectrum, can be adjusted by changing the geometry of the holes and their spacing. So hole arrays can be

made into highly selective filters for sensors that depend on detecting specific colours, or for efficiently extracting monochromatic light from light-emitting diodes (LEDs) and lasers. Indeed, a number of commercial research labs, such as the Panasonic laboratory in Kyoto, Japan, and NEC in Tsukuba, Japan are working on prototypes of plasmon-enhanced devices for displays and telecommunications.

Hole arrays can also be used to channel light into optical devices. In imaging chips for digital cameras, for example, researchers are studying how hole arrays placed on top of individual pixels might help capture incoming light

Plasmons at work

Although plasmonic effects have been known for more than a century, the history of plasmon-based applications began in the early 1970s, when Martin Fleischmann, a chemist at the University of Southampton, UK, and others began to study how light scatters from molecules stuck to a silver surface⁷. Richard Van Duyne, a chemist at Northwestern University in Evanston, Illinois, then discovered this scattering to be enhanced by a seemingly incredible six orders of magnitude⁸.

In today's optimized devices, this enhancement, known as surface-enhanced Raman spectroscopy (SERS), can be several orders of magnitude larger still — strong enough to detect a single molecule⁹. Moreover, SERS has proved very useful in the biochemical and materials sciences by providing information on the chemical composition of molecules at very small concentrations.

SERS is a plasmonic effect: silver nanoparticles act as antennas that take the incoming laser light and, through their surface plasmons, concentrate it. The concentrated light is then scattered by nearby molecules and amplified again by the silver nanoparticles on the way back out. This dual amplification results in a huge overall signal enhancement.

Some applications have reached the market. For example, in

specifically prepared colloids of gold nanoparticles, a clustering of these nanoparticles is triggered by the presence of pregnancy hormones. This leads to a colour change induced by plasmonic effects that has been widely commercialized in pregnancy tests.

The commercialization of SERS has been hampered in many areas by difficulties in achieving highly accurate control over the surface nanostructures. For this reason, researchers are also looking at other sensing techniques such as localized surface plasmon resonance (LSPR). The idea is that, when a surface is covered with nanostructures in the shape of rods or triangles, their plasmonic properties depend strongly on the properties of medium that surrounds them. For example, a solution containing a certain type of molecule has

a refractive index that varies with the concentration of those molecules¹⁰. "These changes to the refractive index lead to measurable changes to the surface plasmon resonance wavelength, which can be observed experimentally," says Stefan Maier from Imperial College London, who studies plasmonic nanostructures and their applications. "The effects can be dramatic." Devices based on LSPR are becoming so sensitive that Van Duyne thinks that they, too, are about to reach the limit of single-molecule detection.

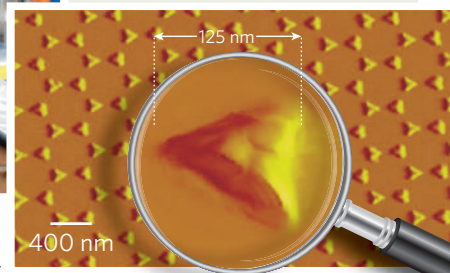
And at Rice University in Houston, Texas, biomedical engineer Naomi Halas is pursuing an optical technique to destroy cancer cells. She hopes to inject cancer patients with gold nanoparticles that will be guided to the tumour by antibodies bound to the particles' surface. Once the

nanoparticles are in place, she can illuminate the area with a low dose of infrared laser light that leaves healthy tissue undamaged, but gets absorbed to create plasmons in the gold. The energy heats up the nanoparticles and kills the cancer cells¹¹.

So far, Halas's cancer therapy has been successful in trials with mice, where she achieved seemingly complete elimination of the tumours. The technology is now in human clinical trials with patients who have head and neck cancers. Halas says the results have been very encouraging so far. "There is no reason one would expect complications from something like this in humans relative to animal trials, because you are using physical mechanisms, heat and light, to induce cell death." Halas is also optimistic that the treatment will be approved for use more quickly than a drug, which can involve difficult and expensive trials and many years to reach the clinic. She says the technique is being considered as a 'device' by the US Food and Drug Administration rather than a drug, which could also accelerate the approval process. **J.H.**



Naomi Halas (centre, above) wants to use plasmons to fight cancer; others use them as sensors (inset) to detect single molecules.



more efficiently, and thus reduce pixel noise and improve camera sensitivity.

Another plasmonic technique for channelling light into a device is to sprinkle its surface with nanoscale particles made of a metal such as gold. These nanoparticles function like an array of tiny antennas: incoming light is taken up by plasmons and then redirected into the device's interior.

Slimming down

From a commercial perspective, perhaps the most promising application of such nano-antennas — or indeed, of hole arrays — is in the improvement of solar cells. Present-day solar cells are made from semiconductors such

as silicon. But to catch as much light as possible from the broadest range of wavelengths, particularly in the red and infrared part of the spectrum, the semiconductor layer has to be relatively thick. "Right now a silicon solar cell is up to 300 micrometres thick," says Albert Polman, a photonics researcher who directs the AMOLF institute in Amsterdam, where he works on improving solar-cell designs. And when cells are being deployed in arrays that cover a rooftop or more, he says, that adds up to a lot of expensive silicon. The price would come down a long way if the silicon was only 1 micrometre thick. "But then you don't catch

the red light because it goes straight through the chip," he says, thus wasting much of the sunlight's available energy. Other solar-cell materials have the same problem.

With plasmonics, however, the problem goes away. In one approach that researchers are exploring, gold nanoparticles on the surface would act as reflectors that focus light into the semiconductor, where absorption efficiency increases with the light concentration. In another scheme, tiny gold nanoantennas could redirect sunlight by 90°, so that it propagates along the semiconductor rather than passing straight through. Either way, the cell

J. C. HULTEEN ET AL. J. PHYS. CHEM. B 103, 3854–3863 (1999)

could get by with a much thinner semiconductor layer.

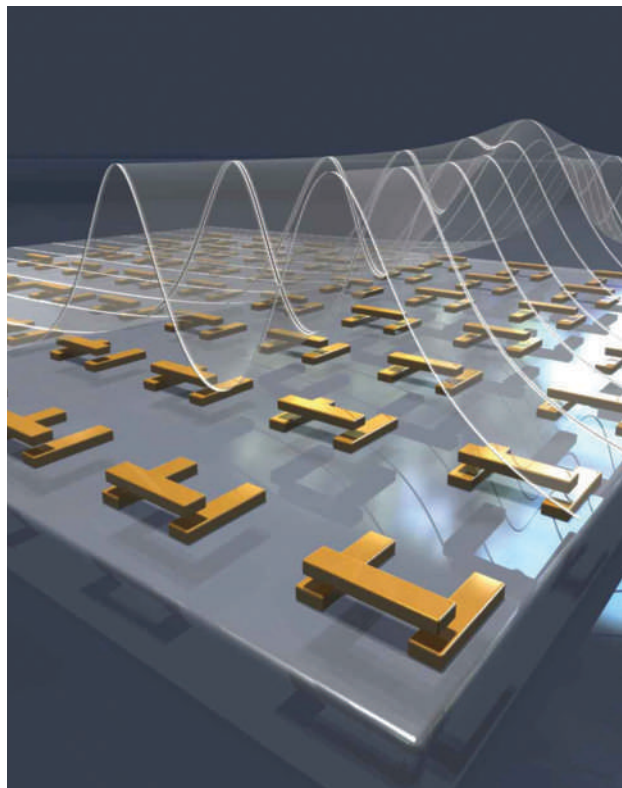
Even as plasmonic techniques are decreasing the cost of the cells, they could also greatly improve the cells' efficiency at extracting the available energy from sunlight — in a field in which even a few percentage points in efficiency improvement are celebrated. Overall, the use of plasmonics could increase the absorption two to five times, says Atwater, who has co-founded Alta Devices in Santa Clara, California, to commercialize such solar cells. For cells made from amorphous silicon, which today have efficiencies of around 10–12%, the predicted enhancements could translate into efficiencies of about 17%. For crystalline silicon cells, which currently have efficiencies around 20%, the new figure could approach the theoretical maximum of 29%. For commercial applications, the remaining challenges include developing workable device designs and fabrication techniques for mass production.

Guiding light

Plasmonics researchers are also grappling with a longer-term challenge: the integration of optics and electronics on a single microchip. The decades-old idea is that, just as a fibre-optic cable can carry much more information than a copper wire, a light beam could, in principle, relay information through the chip on more channels and at a higher speed than conventional integrated circuitry can handle. But the experimental optical devices produced to date have been too large, and have showed rather high losses in the optical signal strength.

"You want to bring the optics closer in size to the transistor," says Polman. And that's the beauty of plasmonics, which can offer optical pathways on virtually the same scale as the silicon structures found in advanced microchips. "Metals can be well integrated with the chip design," says Polman, "so you may be able to distribute light over an integrated circuit by plasmons." Indeed, structures such as silver nanowires² or grooves etched into metal surfaces³ can provide pathways that guide light across a chip in whatever direction the designers might need.

But there is a trade-off as the structures get smaller. If the plasmons are forced to travel through a channel that's too narrow, they start



Plasmon resonance could be used to make very sensitive biochemical sensors (yellow bars). The waves here represent absorption spectra.

"Plasmonics has given photonics the ability to go to the nanoscale."

— Harry Atwater

to leak out from the sides and get lost, says Sergey Bozhevolnyi from the University of Southern Denmark in Odense, who is leading a European research project into integrated plasmonic circuits. Nevertheless, researchers can guide surface plasmons over distances of more than 100 μm , which is roughly a thousand times bigger than the features on a current-generation microchip. This is enough to open rich possibilities for plasmonic nanocircuits, in which light would carry information along complex paths and through many processing steps.

Plasmonic waveguides are particularly promising if the light source — typically a laser — can be incorporated on the chip as well. This has been done with comparatively large lasers, on the order of the wavelength of the laser light. But plasmonics now offers the possibility of doing so at the nanoscale, at lengths much shorter than the wavelength. Rather than amplifying light in a conventional laser cavity, a plasmonic 'spaser' would amplify it with the help of plasmons — the first experimental evidence for such plasmon-based lasing was published in August^{4,5}. To fully integrate these plasmon lasers into standard microcircuitry,

however, researchers will need to find a way to trigger the spasers using standard electrical currents.

In addition to creating light and guiding it across a chip, optical computing will require a way to turn the flow of plasmons on and off at high speeds, so that the flow becomes a series of bits in a digital data stream. Many people have been working on such devices, and a plasmonic modulator based on silicon technology has been realized by Atwater's group. Like a conventional transistor, in which an electric voltage controls a tiny electrical current, the group's device is based on the use of an electric field to control the propagation of surface plasmons through the device⁶. Apart from their small size, compared with conventional optical counterparts, the operation frequency of plasmonic modulators can easily reach tens of terahertz, well above the gigahertz regime of modern computers.

Many roadblocks still remain to the commercialization of such technologies — ranging from the integration with silicon to device issues. "The key thing that keeps coming back are losses in the metals," says Mark Brongersma, a materials scientist at Stanford University in California. However, he adds, smart design of the plasmonic structures could, in principle, reduce losses to acceptable levels.

Plasmonics research has made remarkable progress in the past decade, and researchers are working on pushing our knowledge of plasmons even further, for example to understand the physics very close to the metal surface. Nonetheless, says Atwater, "what has happened in the past seven or eight years is that plasmonics has given to photonics the ability to go to the nanoscale and properly take its place among the nanosciences."

Joerg Heber is a senior editor at Nature Materials.

1. Ebbesen, T. W., Lezec, H. J., Ghaemi, H. F., Thio, T. & Wolff, P. A. *Nature* **391**, 667–669 (1998).
2. Verhagen, E., Spasenović, M., Polman, A. & Kuipers, L. *Phys. Rev. Lett.* **102**, 203904 (2009).
3. Bozhevolnyi, S. I., Volkov, V. S., Devaux, E., Laluet, J.-Y. & Ebbesen, T. W. *Nature* **440**, 508–511 (2006).
4. Noginov, M. A. *et al.* *Nature* **460**, 1110–1112 (2009).
5. Oulton, R. F. *et al.* *Nature* **461**, 629–632 (2009).
6. Dionne, J. A., Diest, K., Sweatlock, L. A. & Atwater, H. A. *Nano Lett.* **9**, 897–902 (2009).
7. Fleischmann, M., Hendra, P. J., McQuillan, A. J. *Chem. Phys. Lett.* **26**, 163–166 (1974).
8. Jeanmaire, D. L. & Van Duyne, R. P. *J. Electroanal. Chem.* **84**, 1–20 (1977).
9. Nie, S. & Emory, S. R. *Science* **275**, 1102–1106 (1997).
10. Anker, J. N. *et al.* *Nature Mater.* **7**, 442–453 (2008).
11. Hirsch, L. R. *et al.* *Proc. Natl Acad. Sci. USA* **100**, 13549–13554 (2003).

CORRESPONDENCE

Sanctions against scientists threaten progress

SIR — Several European countries, including France, the Netherlands and Sweden, are now routinely refusing work visas and study positions in the physical sciences to Iranian scientists and students. This is in response to UN resolution 1737, imposing sanctions on Iran for failing to halt uranium enrichment, and reflects international concern about the potential proliferation of nuclear-weapons technology. The result has been blanket discrimination simply on the basis of nationality.

Similar national-security policies against academics operate in the Middle East. The permit criteria used by Israel's security services for Palestinian postgraduate students are so restrictive that they effectively prohibit entry. Israeli universities have protested, in a letter sent to the defence ministry, that these policies constitute "a gross and harmful intervention by military elements in purely academic considerations" (see <http://go.nature.com/iFljgR>).

The International Council for Science (ICSU) affirms, in its principle of universality in science, that all scientists should have the opportunity to participate in legitimate scientific activities. ICSU's committee on freedom and responsibility in the conduct of science, which I chair, continuously monitors breaches of this principle. We have recently called for the scientific community to commit to opposing all such discrimination (see <http://go.nature.com/2UmM5M>).

Academic institutions should have the responsibility and freedom to select students and staff without political or military interference. If selected individuals are refused entry or a work visa after security screening, the reasons should be made clear to that person.

International collaboration and openness in science

education and research are essential for meeting pressing global challenges. Systematic discrimination against scientists based on nationality is a serious threat to scientific progress.

Bengt Gustafsson Department of Physics and Astronomy, Uppsala University, Box 515, 751 20 Uppsala, Sweden
e-mail: bengt.gustafsson@fysast.uu.se

Measures urgently required to prevent multiple submissions

SIR — A recent experience leads me to believe that defiance of rules against simultaneous submission of papers to different journals may be growing more widespread.

In a thorough review of a submitted paper (not for this journal), I pointed out that the study itself and the organization of the manuscript were below standard; I offered substantive constructive comments and recommended reconsideration after major revision. When the revised manuscript arrived, I made further suggestions for improving its scientific quality.

At this point, and while the manuscript was technically still under consideration by the journal in question, I noticed in a routine online search that it had been published in a different peer-reviewed journal offering rapid publication. Evidently, the authors had submitted the manuscript to the other journal, either simultaneously or after having received the reviewers' comments, without withdrawing it from the first. They had even incorporated some of the comments from my original review.

Cases of duplicate submission are disconcerting for journals and for the scientific community. They seriously violate the principle of disseminating scientific findings with professionalism and integrity. The practice is in breach of the authors' contract to withhold

submission of their manuscript to other journals until the editors have made a formal decision not to publish it.

As the pressure to publish new results rapidly increases and competition becomes ever more intense, editors must define strict reinforcing measures to prevent such violations.

Goudarz Molaei The Connecticut Agricultural Experiment Station, 123 Huntington Street, New Haven, Connecticut 06504, USA
e-mail: goudarz.molaei@ct.gov

Nature journals forbid duplicate submission: <http://go.nature.com/dthpdU>

Caution with claims that a species has been rediscovered

SIR — We welcome the recent announcement by the conservation partnership BirdLife International that they have launched a "global bid to try to confirm the continued existence of 47 species of bird that have not been seen for up to 184 years" (see <http://go.nature.com/6Hc2Cn>). But there are pitfalls, as the recent history of 'rediscoveries' has shown.

One of the species on BirdLife's target list is the ivory-billed woodpecker (*Campephilus principalis*), a bird that was prematurely alleged to have been rediscovered in 2005. This seemingly improbable reappearance provoked intense debate within the scientific community about the veracity of claimed sightings and, more generally, about what represents sufficient proof of continued existence (or extinction). Accusations of 'faith-based' ornithology resulted, increasing scepticism among politicians and policy-makers that conservation organizations are often too willing to put public relations before scientific rigour.

Many rediscoveries in the developing world are made by

individuals or organizations from Europe or the United States, or are a direct result of Western-backed expeditions or initiatives. This wrongly reinforces the impression that only Western scientists are competent to find and save threatened species. In addition, high-profile rediscoveries can create an unexpected imperative for immediate action by hard-pressed national conservation organizations.

The international conservation community often seems to want it both ways, being unwilling to declare a species extinct but enthusiastically proclaiming the rediscovery of an 'extinct' species. This ambiguity is understandable — high biogeographic uncertainty can be generated both by the IUCN Red List requirements for 'exhaustive surveys' before a species is officially declared extinct, and by frequent taxonomic revisions that propel rarely seen subspecies to full species status. Rediscoveries are only meaningful if backed up by a self-sustaining population. Otherwise, conservationists are merely engaged in the sad task of documenting the prolonged demise of yet another species.

The genuine rediscovery of 'lost' species is a newsworthy event that helps bolster the pioneering, field-based credentials of conservation and draws attention to new sites worthy of increased protection. The combination of technology and improved access makes finding these species easier than ever. The real challenge is how to present rediscoveries to the public in a way that reflects their conservation significance and that will best encourage the support of future conservation efforts.

Richard J. Ladle, Paul Jepson Oxford University Centre for the Environment, Oxford OX1 3QY, UK
e-mail: richard.ladle@ouce.ox.ac.uk
Steve Jennings Oxfam GB, South Asia RMC (Delhi), New Delhi, India
Ana C. M. Malhado Department of Agricultural and Environmental Engineering, Federal University of Viçosa, Viçosa, Minas Gerais, Brazil

OPINION

An agenda for personalized medicine

Pauline C. Ng, Sarah S. Murray, Samuel Levy and J. Craig Venter find differences in results from two direct-to-consumer genetics-testing companies. They therefore give nine recommendations to improve predictions.

More than 1,000 DNA variants associated with diseases and traits have been identified^{1,2}. Direct-to-consumer (DTC) companies are harnessing these discoveries by offering DNA tests that provide insights into personal genetic traits and disease risks. Genetic testing can improve lifestyle choices and increase preventive screening³. However, understanding of the genetic contribution to human disease is far from complete.

There is debate in the genetics community as to the usefulness of DTC testing. Therefore, we compared results from two DTC companies (test kits provided by genomics companies 23andMe in Mountain View, California, and Navigenics in Foster City, California) on 13 diseases for 5 individuals. Despite this limited data set we find potential implications for personalized medicine. Here we provide recommendations to improve predictions and support the continued growth of this nascent industry.

DTC genome scans are easy to get. Users order tests online, provide saliva or a cheek swab, and within a few weeks 500,000–1,000,000 of their DNA variant markers are scanned. The service provider then calculates a set of disease risks based on the customer's specific combination of markers, and presents the results to the user online (see graphic).

The accuracy of DTC genome-scan tests has been questioned. It is our assessment that the accuracy of the raw data is high. We found that the genotypes, or particular DNA bases observed, of an individual's markers from 23andMe and Navigenics agreed more than 99.7% of the time. This is similar to accuracies reported by the genotyping companies.

Two other major concerns are whether the predicted disease risks have any clinical validity, and how well a genetic variant correlates with a specific disease or condition⁴. A few individuals have alluded to getting different predictions from different DTC companies for the same disease^{5,6}. We compared the consistency of disease-risk predictions between the two DTC companies to see where differences may arise (see Table 1).

Both companies report absolute risk, which is the probability that an individual will develop a disease. Absolute risk is derived from two parameters: 'relative risk' and 'average population disease risk'. Relative risk is modelled from an individual's genetics. Average population disease risk varies depending on how one defines the population. For example, Navigenics distinguishes population disease risk between men and women (for example, men are more likely to have heart attacks than women), whereas 23andMe primarily takes into account age (for example, incidence of rheumatoid arthritis increases with age). This ambiguity in the definition of a 'population' underscores the caution one must exercise when interpreting absolute risk results.

Even after we removed the average population risk variable we still found that only two-thirds of relative risk predictions qualitatively agree between 23andMe and Navigenics when averaged across our five individuals (see Table 1). Certain diseases have better prediction agreement than others. For four diseases, the predictions between the two companies completely agree for all individuals. In contrast, for seven diseases, 50% or less of the predictions

SUMMARY

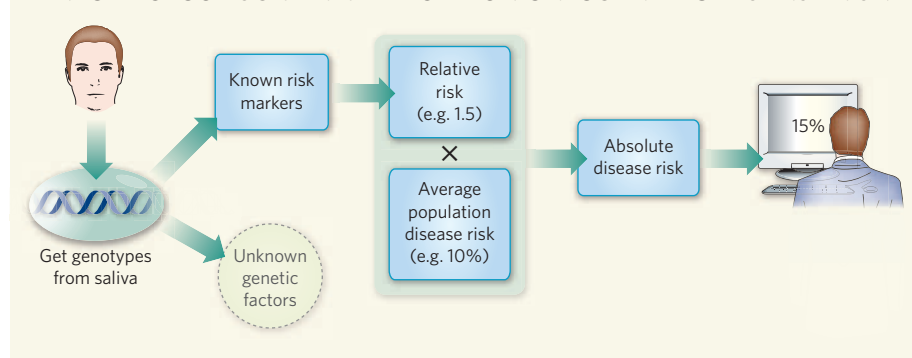
- For seven diseases, 50% or less of the predictions of two companies agreed across five individuals
- Companies should communicate high risks better and test for drug response markers
- Community should study markers in all ethnicities and look at behaviour after tests

agree between the two companies across the individuals.

A major contributor to the discrepancies in disease-risk predictions is the set of markers that each service chooses to use in calculating relative risk. Risk markers are determined from genome-wide association studies, which survey hundreds of thousands or millions of markers across control and disease patients⁷. Each marker has different possible alleles. Alleles that occur more frequently in disease patients are designated as risk alleles and have odds ratios greater than 1. For example, in Alzheimer's disease patients, 38% of *ApoE* alleles are the *ApoE4* risk allele; this allele's frequency is only 14% in normal controls⁸. The odds ratio for the *ApoE4* risk allele is 3.7 (odds of exposure in cases, divided by odds of exposure in controls is $(0.38/0.62)/(0.14/0.86)$). The greater the frequency disparity between disease patients and normal controls, the higher the odds ratio associated with the allele. Conversely, alleles conferring protection against disease are observed less frequently in disease patients and have odds ratios less than 1.

DTC companies harness the same publicly available research to decide which markers to include, and for the most part, could use the same or similar markers. Yet no disease has an identical set of markers between the two DTC companies because each company has its own criteria for accepting a genome-wide association result into its relative risk calculation^{9,10}. Some markers are used by both companies for a particular disease. For identical markers and correlated markers, the odds ratios are similar between the two DTC companies ($r=0.98$ for identical markers; $r=0.89$ for correlated markers). In other words, once DTC companies agree that a marker is predictive of disease, they tend to agree on its genetic contribution to disease predictions.

DIRECT-TO-CONSUMER METHOD FOR CALCULATING DISEASE RISK



What gives two similar DTC predictions? Coeliac disease is one condition for which predictions agreed between the two companies for all five individuals in our analysis. For coeliac disease, both companies have one strong-effect marker with a high odds ratio in common; Navigenics also reports on seven additional markers that 23andMe does not use. Thus, the number of markers in common does not necessarily correlate with better prediction agreement. However, the one strong-effect marker in common between both companies occurs in >90% of people with coeliac disease¹¹ and its risk allele has an odds ratio of 7 (ref. 12). The seven markers unique to Navigenics have modest effects¹² and therefore do not affect the overall relative risk prediction for this disease as much. Generally predictions tended to agree when there was consensus on the strong-effect markers for a disease.

When the DTC companies did not use the same strong-effect markers, we saw large differences in prediction. A clear example is the predicted disease risk for psoriasis. In one individual, 23andMe reports a relative risk of 4.02, whereas Navigenics reports a relative risk of 1.25, more than a threefold difference. The difference is attributable to a marker unique to 23andMe whose risk allele has an odds ratio of 2.8 (ref. 13). This marker is not included in the Navigenics analysis because the result does not seem to pass Navigenics' publication requirements for marker inclusion.

Another concern is the use of markers that have uncertain odds ratios estimates. A marker for type 2 diabetes that Navigenics uses has the highest odds ratio among all of Navigenics' type 2 diabetes markers as reported in the literature. It therefore makes the strongest contribution to the overall disease prediction. However, Navigenics warns that the marker's effect is statistically insignificant and may not contribute to disease. The average consumer is unlikely to appreciate the significance, or lack thereof, of this result.

These findings lead us to propose the following recommendations for a personalized medicine research agenda.

Company recommendations

Report the genetic contribution for the markers tested. Currently, the markers that have been discovered by genome-wide association studies do not explain the majority of the genetic heritability of disease. For example,

TABLE 1: PREDICTIONS FOR DISEASE RELATIVE RISKS FOR FIVE INDIVIDUALS

Disease	Female A	Female B	Female C	Male D	Male E
Breast cancer	↑↑	↑↑	↓↓		
Coeliac disease	↓↓	↓↓	↓↓	↓↓	↓↓
Colon cancer	==	==	=↓	↑↑	=↓
Crohn's disease	↓↑	↓↑	↓↓	↓↓	↓=
Heart attack	↓↓	=↓	=↓	=↓	↑↑
Lupus	↑↓	↓↓	↓↓	↑=	↑=
Macular degeneration	↓↓	↓↓	↑=	↓↓	↓↓
Multiple sclerosis	↑↑		↓↓	↓↓	↓↓
Prostate cancer				↑↑	↓↑
Psoriasis	↓↑		↑↓	↑↑	↓↓
Restless legs syndrome	=↓	↑↑	↓=	↓↑	↑↑
Rheumatoid arthritis	↑↑	↑↑	↓↓	↓↓	↑↑
Type 2 diabetes	↓↓	=↓	↓↓	↑↓	=↓

↑ increased risk (RR > 1.05), ↓ decreased risk (relative risk (RR) < 0.95), = average risk (0.95 ≤ RR ≤ 1.05). First prediction is from 23andMe; second prediction is from Navigenics. Different predictions are highlighted in beige.

current literature indicates that approximately 60–65% of the heritability of coeliac disease is still unaccounted for. Therefore, the marker set used to screen for disease can miss unknown genetic factors, leading to false negatives. We recommend that DTC companies report the proportion of the genetic contribution of a disease that can be attributed to the markers used in their test, and the proportion of the genetic contribution that is still unknown. This is different from reporting the genetic contribution versus the environmental contribution, which DTC companies emphasize on their websites.

Focus on high-risk predictions. Most of the diseases predicted in the DTC reports imply only a modest risk compared with the average population (approximately 80% of reported relative risks lie between 0.5 and 1.5). We recommend that DTC companies structure their communications with users around diseases and traits that have high-risk predictions. Customers could focus their lifestyle changes based on these. However, if there is a low-risk prediction for disease, a sense of security should not be assumed because much of the genetic contribution to disease risk has yet to be understood.

Directly genotype risk markers. If the risk marker in the published literature is not directly assayed by the DTC company, DTC companies currently use linkage disequilibrium (the non-random association of alleles) to choose a surrogate risk marker. We

estimate that 1% of markers in an individual will violate the assumption of perfect linkage disequilibrium, and in these cases, using a surrogate marker would be misleading. Although this percentage seems small, hundreds of markers are tested so there is likely to be at least one error. Instead, direct genotyping of the disease-associated marker would improve accuracy of risk genotypes. Some DTC companies already target specific key markers not on the whole genome arrays or specific markers that have failed on the whole genome arrays¹⁰.

Test pharmacogenomic markers.

An estimated 100,000 people die annually in the United States from adverse drug reactions^{14,15}. Although few drugs are labelled to require or recommend genetic testing, consumers could find specific variants useful. Variants in drug metabolism genes or recommended for test-

ing by drug labels are informative and could greatly affect an individual's treatment. Examples include variants affecting the efficacy of clopidogrel (used to reduce the risk of stroke or heart attack) or tamoxifen (used to treat breast cancer)¹⁶. Most of the DTC companies are testing for some pharmacogenomic markers^{17,18}; we encourage inclusion of as many of these markers as possible.

Agree on strong-effect markers. DTC companies have agreed to use clinically validated markers for prediction, but not necessarily the same markers or number of markers¹⁹. This lack of consensus leads to inconsistent results between DTC companies. As studies are replicated, the number of markers and better estimates of their odds ratios should converge so that there is consensus to include a marker. Because these studies will take time, a stopgap solution is for DTC companies to agree on using a core set of strong-effect markers to achieve better prediction consensus and consistent reporting to the consumer.

Community recommendations

Monitor behavioural outcomes. One of the fundamental questions with DTC tests is whether they modify consumers' behaviour long term, and hence benefit lifestyle and health²⁰. More public studies need to be funded to monitor behaviour resulting from DTC testing to identify the best strategies for using personal genomic data to improve an individual's health. Studies are currently under

way, and applying the findings will be central to the success of DTC genome tests and credibility of the field²⁰.

Carry out prospective studies. Agreement on risk predictions by DTC companies does not necessarily imply that the predictions are accurate or meaningful, and at this point in time, we cannot determine who has the 'best' predictions. To effectively assess the clinical validity of these genetic tests the community needs more prospective studies with tens or hundreds of thousands of individuals that measure the predictive value of known markers^{21–23}. Such studies are useful because they consider risk markers simultaneously, measure the interaction between different markers and do not assume a risk model. It may be practical to prioritize common diseases with significant health impact because of the large numbers of individuals and the expense associated with prospective studies.

Replicate associated markers in other ethnicities. Genome-wide association studies have been conducted primarily on populations with European ancestry¹. Disease-associated markers may not transfer from one population to another — allele frequencies or linkage disequilibrium patterns may differ^{1,24}. Therefore, we strongly recommend the validation of these markers and the surrounding pattern of genetic variation in other ethnicities.

Sequence rather than genotype. Eventually, sequencing an individual's genome will become economically feasible. Sequencing has an advantage over genotyping because it captures the full spectrum of an individual's variation and determines, rather than infers, a higher resolution of variants. However, identification of variants should not be confused with their interpretation, and pinpointing the causative disease variant will still be challenging²⁵. Our ability to identify variants from comprehensive sequence data will far outstrip our ability to characterize their biological effect. However, accurate and complete reporting is a necessary predecessor to a precise functional understanding of genomic data for the consumer. ■

Pauline C. Ng and **J. Craig Venter** are at the J. Craig Venter Institute, Science Center Drive, San Diego, California 92121, USA. **Sarah S. Murray** is at Scripps Genomic Medicine, Scripps Translational Science Institute, North Torrey Pines Court, La Jolla, California 92037, USA. **Samuel Levy** is at the J. Craig Venter Institute, Medical Center Drive, Rockville, Maryland 20850, USA.
e-mail: png@jcv.org

1. Frazer, K. A., Murray, S. S., Schork, N. J. & Topol, E. J. *Nature Rev. Genet.* **10**, 241–251 (2009).
2. Mailman, M. D. et al. *Nature Genet.* **39**, 1181–1186 (2007).
3. Heshka, J. T., Palleschi, C., Howley, H., Wilson, B. & Wells, P. S. *Genet. Med.* **10**, 19–32 (2008).
4. Hogarth, S., Javitt, G. & Melzer, D. *Annu. Rev. Genomics Hum. Genet.* **9**, 161–182 (2008).
5. Duncan, D. E. *Experimental Man: What One Man's Body*

6. *Reveals About His Future, Your Health, and Our Toxic World* (John Wiley & Sons, 2009).
7. Singer, E. *Technology Review* (MIT, 15 June, 2009); available at www.technologyreview.com/blog/editors/23680
8. Topol, E. J., Murray, S. S. & Frazer, K. A. *J. Am. Med. Assoc.* **298**, 218–221 (2007).
9. Bertram, L., McQueen, M. B., Mullin, K., Blacker, D. & Tanzi, R. E. *Nature Genet.* **39**, 17–23 (2007).
10. Hsu, A. & Naughton, B. (eds) *Guidelines on Vetting Genetic Associations* (23andMe; 2008); available at https://23andme.com/res/4400/pdf/23-03_Vetting_Genetic_Associations.pdf
11. *Applying Preventive Genomic Medicine in Clinical Practice* (Navigenics, 2008); available at www.navigenics.com/static/pdf/physician/physician-whitepaper.pdf
12. Green, P. H. R. & Jabri, B. *Lancet* **362**, 383–391 (2003).
13. Hunt, K. A. et al. *Nature Genet.* **40**, 395–402 (2008).
14. Liu, Y. et al. *PLoS Genet.* **4**, e1000041 (2008).
15. Lazarou, J., Pomeranz, B. H. & Corey, P. N. *J. Am. Med. Assoc.* **279**, 1200–1205 (1998).
16. Pirmohamed, M. et al. *Br. Med. J.* **329**, 15–19 (2004).
17. Sangkuhl, K., Berlin, D. S., Altman, R. B. & Klein, T. E. *Drug Metab. Rev.* **40**, 539–551 (2008).
18. www.23andme.com/health/all
19. www.pathway.com/more_info/drug_responses
20. Personalized Medicine Coalition *Personal Genomics and Industry Standards: Scientific Validity* (2008); available at <http://cancercontrol.cancer.gov/od/phg/docs/PMCSiValid.pdf>
21. Prainsack, B. et al. *Nature* **456**, 34–35 (2008).
22. Aulchenko, Y. S. et al. *Nature Genet.* **41**, 47–55 (2009).
23. Lyssenko, V. et al. *N. Engl. J. Med.* **359**, 2220–2232 (2008).
24. Meigs, J. B. et al. *N. Engl. J. Med.* **359**, 2208–2219 (2008).
25. Adeyemo, A. & Rotimi, C. *Public Health Genomics* doi: 10.1159/000218711 (2009).
26. Easton, D. F. et al. *Am. J. Hum. Genet.* **81**, 873–883 (2007).

Supplementary information accompanies this article online. The authors have disclosed a statement of conflicting interests that is available online.

See Editorial page 697, and online at <http://go.nature.com/VqPUE2>.

Let's celebrate human genetic diversity

Science is finding evidence of genetic diversity among groups of people as well as among individuals. This discovery should be embraced, not feared, say **Bruce T. Lahn** and **Lanny Ebenstein**.

A growing body of data is revealing the nature of human genetic diversity at increasingly finer resolution^{1,2}. It is now recognized that despite the high degree of genetic similarities that bind humanity together as a species, considerable diversity exists at both individual and group levels (see box, page 728). The biological significance of these variations remains to be explored fully. But enough evidence has come to the fore to warrant the question: what if scientific data ultimately demonstrate that genetically based biological variation exists at non-trivial levels not only among individuals but also among groups? In our view, the scientific community and society at large are ill-prepared for such a possibility. We need a moral response to this question that is robust

irrespective of what research uncovers about human diversity. Here, we argue for the moral position that genetic diversity, from within or among groups, should be embraced and celebrated as one of humanity's chief assets.

The current moral position is a sort of 'biological egalitarianism'. This dominant position emerged in recent decades largely to correct grave historical injustices, including genocide, that were committed with the support of pseudoscientific understandings of group diversity. The racial-hygiene theory promoted by German geneticists Fritz Lenz, Eugene Fischer and others during the Nazi era is one notorious example of such pseudoscience. Biological egalitarianism is the view that no or almost no meaningful genetically

based biological differences exist among human groups, with the exception of a few superficial traits such as skin colour³. Proponents of this view seem to hope that, by promoting biological sameness, discrimination against groups or individuals will become groundless.

We believe that this position, although well-intentioned, is illogical and even dangerous, as it implies that if significant group diversity were established, discrimination might thereby be justified. We reject this position. Equality of opportunity and respect for human dignity should be humankind's common aspirations, notwithstanding human differences no matter how big or small. We also think that biological egalitarianism may not remain viable in light of the growing body of empirical data (see box).



Will we soon cherish genetic diversity as we now do cultural diversity?

Many people may acknowledge the possibility of genetic diversity at the group level, but see it as a threat to social cohesion. Some scholars have even called for a halt to research into the topic or sensitive aspects of it, because of potential misuse of the information⁴. Others will ask: if information on group diversity can be misused, why not just focus on individual differences and ignore any group variation?

We strongly affirm that society must guard vigilantly against any misuse of genetic information, but we also believe that the best defence is to take a positive attitude towards diversity, including that at the group level. We argue for our position from two perspectives: first, that the understanding of group diversity can benefit research and medicine, and second, that human genetic diversity as a whole, including group diversity, greatly enriches our species.

For scientists to understand human genetic diversity in its totality, group diversity cannot be shunned. It is an integral and meaningful component of overall human diversity. For example, the International HapMap Project, which examines genetic diversity in several

hundred individuals, has revealed clear genetic differentiation among the geographic groups represented by those individuals. More importantly, studies increasingly indicate that understanding genetic diversity at the group level can shed light on human evolution, the nature and acquisition of many human traits, including disease conditions, and how genetic and environmental factors interact to produce biological outcomes^{1,2,5,6}. Thus, to ignore group diversity is to do poor science.

Neither can many medical applications of genetics safely ignore group diversity. It can facilitate the mapping of disease genes and lead to improved treatments^{7,8}. For instance, groups can differ markedly in their ability to metabolize certain anticancer drugs⁸. Examining the potential genetic contributions to these differences may illuminate how genes regulate drug metabolism and allow for more effective treatment. The ultimate goal of medical intervention may be personalized medicine (see page 724), whereby treatment is tailored to the genetic make-up of each individual, but this will remain a distant ideal for years to come. For now, to intentionally overlook the influence of group diversity on disease susceptibilities and treatment outcomes is to practise poor medicine.

In addition to the above arguments, there is a much larger reason to embrace human diversity in all its forms, in our view. Humanity's genetic diversity — small or large, within or among groups — is a resource for, rather than a detriment to, creating a more fulfilling and prosperous society. Just as people have come over time to cherish cultural diversity, so we hope that

attitudes will warm towards genetic diversity.

In the natural world, genetic diversity is a source of evolutionary resilience and adaptability. It buffers against changing environments and allows species to occupy broader and more fluid ecological niches. Even for a single individual, differences between its two copies of the genome can often lead to higher fitness. Indeed, sexual reproduction is thought to have evolved as a way for species to take advantage of genetic diversity. Consequently, the loss of a species' diversity often threatens its long-term survival. The susceptibility of agro-monoculture to sudden disease outbreaks or climate changes is just one example.

In humans, genetic diversity may be particularly beneficial at a social or cultural level. Humans are uniquely complex in the social structures they form. Although genetic diversity may not be a prerequisite for social complexity, the former can foster the latter by allowing individuals with different tastes and abilities to make professional and personal choices that they enjoy and in which they are productive, thereby leading to personal fulfilment and contributing to a more prosperous society. Arguably, the United States is one of the most innovative, successful and culturally vibrant countries in the world. It excels in numerous and wide-ranging areas, such as art, sport, business, science and political and economic thought. We believe that this is due in part to the nation's diversity, both cultural and genetic, and to a liberal environment that allows individuals to pursue their unique and varied potentials.

Group differentiation, by furthering this

SUMMARY

- Promoting biological sameness in humans is illogical, even dangerous
- To ignore the possibility of group diversity is to do poor science and poor medicine
- A robust moral position is one that embraces this diversity as among humanity's great assets

diversity, adds to the total depth and breadth of human potential. In our view, the 2008 Olympic Games was a beautiful showcase of human diversity. Diversity at the individual level was evident from the wide-ranging physical attributes associated with different sports. Often athletes from different geographic areas also seem to excel at certain sports. Many of these differences may of course be explained by cultural and environmental influences, but should genetic variation contribute in any way to regional athletic ability, it would be hard not to see group diversity as a great asset of our species.

Discussions of human genetic diversity inevitably touch on many sensitive issues. We therefore provide the following caveats to minimize misinterpretations of our position. First, the recognition that genetic diversity can contribute to variation in biological traits by no means diminishes the role of the environment in influencing many of these traits. Arguments for improving the well-being of individuals and groups through environmental approaches such as better nutrition, education, career opportunities and medical treatment lose none of their strength when embracing genetic diversity. Second, acknowledging differentiation among groups does not reduce the importance of diversity within groups, in which most human diversity seems to lie. Third, although we firmly believe that diversity is beneficial overall, we acknowledge that it might not always be so. For example, genetic diversity can lead to higher disease susceptibilities in some individuals or groups. We nevertheless believe that any downside of genetic diversity, including at the group level, does not detract from its overall benefit to our species.

It is also important to recognize that humanity is diverse in its diversity — which is to say that genetic diversity contributes to variation across numerous physical, physiological and cognitive domains. How individuals or groups fare in one domain can be largely independent of how they fare in others. For example, although IQ is a useful metric of some aspects of intelligence and it is partly heritable, it is far from a complete measure of total mental capacity. Therefore, acceptance of human genetic diversity in its totality necessarily leads to the rejection of unidimensional rankings of the capacity of human individuals or groups. If anything, the study of genetics is taking us towards an ever greater appreciation of the multidimensional nature of human potential.

Genetic diversity is a strength not a weakness of humanity. It is time to acknowledge, embrace

Emerging understanding of human genetic diversity

Genetic diversity is the differences in DNA sequence among members of a species. It is present in all species owing to the interplay of mutation, genetic drift, selection and population structure. When a species is reproductively isolated into multiple groups by geography or other means, the groups differentiate over time in their average genetic make-up.

Anatomically modern humans first appeared in eastern Africa about 200,000 years ago. Some members migrated out of Africa by 50,000 years ago to populate Asia, Australia, Europe and eventually the Americas⁹. During this period, geographic barriers separated humanity into several major groups, largely along continental lines, which greatly reduced gene flow among them. Geographic and cultural barriers also existed within major groups, although to lesser degrees.

This history of human demography, along with selection, has resulted in complex patterns of genetic diversity. The basic unit of this diversity is polymorphisms — specific sites in the genome that exist in multiple variant forms (or alleles). Many polymorphisms involve just one or a few nucleotides, but some may involve

large segments of genetic material². The presence of polymorphisms leads to genetic diversity at the individual level such that no two people's DNA is the same, except identical twins. The alleles of some polymorphisms are also found in significantly different frequencies among geographic groups¹⁵. An extreme example is the pigmentation gene *SLC24A5*. An allele of *SLC24A5* that contributes to light pigmentation is present in almost all Europeans but is nearly absent in east Asians and Africans¹⁰.

Given these geographically differentiated polymorphisms, it is possible to group humans on the basis of their genetic make-up. Such grouping largely confirms historical separation of global populations by geography⁵. Indeed, a person's major geographic group identity can be assigned with near certainty on the basis of his or her DNA alone (now an accepted practice in forensics). There is growing evidence that some of the geographically differentiated polymorphisms are functional, meaning that they can lead to different biological outcomes (just how many is the subject of ongoing research). These polymorphisms

can affect traits such as pigmentation, dietary adaptation and pathogen resistance (where evidence is rather convincing)^{10–12}, and metabolism, physical development and brain biology (where evidence is more preliminary)^{6,8,13,14}.

For most biological traits, genetically based differentiation among groups is probably negligible compared with the variation within the group. For other traits, such as pigmentation and lactose intolerance, differences among groups are so substantial that the trait displays an inter-group difference that is non-trivial compared with the variance within groups, and the extreme end of a trait may be significantly over-represented in a group.

Several studies have shown that many genes in the human genome may have undergone recent episodes of positive selection — that is, selection for advantageous biological traits⁶. This is contrary to the position advocated by some scholars that humans effectively stopped evolving 50,000–40,000 years ago¹⁵. In general, positive selection can increase the prevalence of functional polymorphisms and create geographic differentiation of allele frequencies. **B.T.L. & L.E.**

"Geographic group identity can be assigned on DNA alone."

and celebrate this strength. There is nothing scientifically improbable or morally reprehensible in the position that people, including groups of people, can be genetically diverse. Those who deny or even condemn human diversity adopt a stance that is both factually doubtful and morally precarious. On the whole, humanity has been and will be stronger, not despite our differences, but because of them. ■

Bruce T. Lahn is in the Department of Human Genetics, University of Chicago, Illinois. **Lanny Ebenstein** is in the Department of Economics, University of California at Santa Barbara, California. e-mail: blahn@bsd.uchicago.edu

1. Frazer, K. A. *et al.* *Nature* **449**, 851–861 (2007).
2. Redon, R. *et al.* *Nature* **444**, 444–454 (2006).
3. Kamin, L. J., Lewontin, R. C. & Rose, S. *Not in Our Genes: Biology, Ideology, and Human Nature* (Pantheon, 1984).
4. Rose, S. *Nature* **457**, 786–788 (2009).
5. Li, J. Z. *et al.* *Science* **319**, 1100–1104 (2008).
6. Pickrell, J. K. *et al.* *Genome Res.* **19**, 826–837 (2009).
7. Zhu, X., Tang, H. & Risch, N. *Adv. Genet.* **60**, 547–569 (2008).
8. Phan, V. H. *et al.* *Expert Opin. Drug Metab. Toxicol.* **5**, 243–257 (2009).
9. Peregrine, P. N., Peiros, L. & Feldman, M. (eds) *Ancient Human Migrations: A Multidisciplinary Approach* (Univ. Utah Press, 2009).
10. Lamason, R. L. *et al.* *Science* **310**, 1782–1786 (2005).
11. Tishkoff, S. A. *et al.* *Nature Genet.* **39**, 31–40 (2007).
12. Hamblin, M. T. & Di Rienzo, A. *Am. J. Hum. Genet.* **66**, 1669–1679 (2000).
13. Fujimoto, A. *et al.* *Hum. Mol. Genet.* **17**, 835–843 (2008).
14. de Quervain, D. J.-F. *et al.* *Nature Neurosci.* **10**, 1137–1139 (2007).
15. Gould, S. J. *The Structure of Evolutionary Theory* (Belknap Press, 2002).

Have your say at <http://go.nature.com/176Rzs>. See also Editorial, page 697, and online at <http://go.nature.com/VqPUE2>.

BOOKS & ARTS

Winning the arguments on Capitol Hill

Harold Varmus enjoys a guide to the inner workings of the US Congress by legislator Henry Waxman.

**The Waxman Report:
How Congress Really Works**
by Henry Waxman with Joshua Green
Twelve: 2009. 256 pp. \$24.99

Most US scientists who are politically engaged on behalf on their profession have one objective: to enhance the budgets of their funding agencies. Their heroes are supportive advocates and congressional appropriators. But the rest of what Congress does may seem irrelevant, irrational or even mysterious.

In his first book, Henry Waxman — a Democrat, a member of the US House of Representatives since 1975, and one of the most accomplished legislators of our time — gives a useful corrective, focusing on policy and oversight, not just the money. *The Waxman Report* is a welcome guide for those who wish to learn more about the complex intersections of science and government, as the author describes his legislative fights against tobacco, HIV/AIDS and the use of steroids in sports; and his advocacy of food nutrition labelling, clean air and drugs for rare diseases.

Waxman represents the district that includes Hollywood, California, but he would not be called glamorous in appearance or style. He does, however, share other traits with his district's most famous industry — an aptitude for dramatic staging, an appetite for intriguing strategies and a recognition of star power. Some of his most stirring moments have come when using his hearing room as a stage to assemble powerful figures — from captains of industry to sports heroes — to expose deceptions that threaten public health or the environment. In this sense, he more closely resembles a morally driven film director than a committee chairman.

Waxman's legislative successes have often depended on understanding the importance of public support, shrewdly assessing how to get it, and effectively transmitting the message to key people. In 1994, he knew that Congress would not give the US Food and Drug Administration (FDA) any regulatory authority over tobacco products. But he also knew that he could



Henry Waxman (below left) made tobacco executives testify in Congress in 1994 to sway public opinion.

move public opinion in the right direction by bringing the heads of tobacco companies to testify before Congress and then asking embarrassing questions. That famous hearing — and the iconic photograph of the mass swearing-in of chief executives — helped to build public support for the extended powers over tobacco products recently granted to the FDA by Congress.

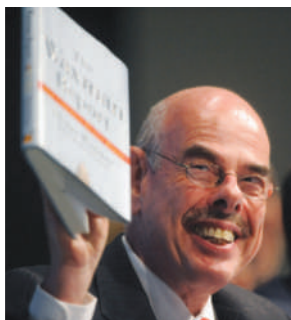
Some years earlier, when Waxman's bill to provide tax benefits for companies that made drugs for rare illnesses was threatened with a Senate defeat or a presidential veto, he asked friends in Hollywood to produce a television show that dramatized the problem, and asked others to lobby President Ronald Reagan at a holiday party. In this way, the Orphan Drug Act became law in 1983. He also praises Edward Kennedy's naming of one of the first major pieces of HIV/AIDS legislation in 1989 after Ryan White, a young patient with haemophilia who had been infected by a blood transfusion and who happened to live in a midwestern state represented by a senator whose vote was crucial. Even homophobic legislators were unlikely to oppose the Ryan White CARE Act.

Gimmicks, of course, do not work on their own. Waxman's successes have required a passion for progressive policy, patience, persistence

and a deep knowledge of his subjects — traits that are all too rarely displayed in Congress these days. Waxman preaches a seemingly naive optimism, noting that “no matter how gloomy the outlook or fearsome the opposition ... landmark legislation can be attained through organization, skill, and hard work”. In fact, his victories have often depended on savviness as well as on industry. As a proponent of compromise with his opponents, he writes of the virtue of being open to “unlikely alliances”. And as a tactician, he notes that whereas big issues generate noise, they often have little effect on ordinary people's lives. Smaller issues such as food labelling, he explains, may “fly under the radar, but ... have a revolutionary impact”.

Waxman is tough and pragmatic, as well as clever and idealistic. He defends his provision of campaign funds to fellow Democrats who might later support his bills, saying that it is “useful to think of money as a political fact of life”. He speaks frankly about his opponents' faults and about his own occasional missteps, such as the day he yielded to an unfortunate compromise on the labelling of dietary supplements. And he recounts how he has made use of procedural tactics to achieve his ends, such as bringing the legislative process on a colleague's weak ‘clean air’ measure to a near-standstill.

Despite differences in social background, Waxman has much in common with the late Edward Kennedy. Recently eulogized as perhaps the most effective senator of the modern era, Kennedy was heir to a familial political tradition, entering the Senate with ease at a young age despite little experience. By contrast,



**“Landmark legislation
can be attained
through organization,
skill and hard work.”**
— Henry Waxman

Waxman was raised by a struggling Jewish family working its way out of the Depression in Los Angeles. After fighting his way, with a few lucky breaks and a law degree, to the California State Assembly in 1968, he positioned himself in 1974 for a newly created seat in the House of Representatives and won a subcommittee chairmanship only five years later.

In their respective halls of Congress, Kennedy and Waxman became similarly known as masters of the legislative process, combining liberal political ideals with a willingness to work with opponents to get things done. During long

careers, both have produced remarkable legislative records in domains in which science is important, including health care and regulatory policy — yet without ever serving on those all-powerful appropriations committees. In this slim volume, we learn how Waxman did it. ■

Harold Varmus is a former director of the US National Institutes of Health and author of *The Art and Politics of Science*. He is president of the Memorial Sloan-Kettering Cancer Center, New York, and a co-chair of President Obama's Council of Advisors on Science and Technology. e-mail: varmus@mskcc.org

inside China was the founding in 1995 of the Tsinghua Bulletin Board System (BBS), which was started by students at the computer-science department of Tsinghua University, where I was an undergraduate. Even today, with the prevalence of text messaging, blogs, YouTube and Twitter, the BBS continues to be a widely used online platform in China, and its underlying technology has progressed from dial-up connections to broadband networks.

Although filled with vivid anecdotes, this book is an academic publication. Its storytelling is punctuated by jargon and scholarly narratives, including numerous academic references. Nonetheless, it is a valuable information resource. Yang's analysis covers a broad canvas and includes many statistics. The investigation into the business side of online activism will particularly fascinate many readers. Online viewings surely translate into money, and manufactured online contention generates lots of viewings. Some businesses, including art dealers, present items as 'banned in China' to promote their wares. Also a reality are competitive tactics, such as the '50 cents party' — people who are paid 50 cents an item for posting prescribed messages at online forums.

Governmental control of content is the elephant in the room. The mechanisms for restricting content flow into China and for controlling domestic Internet content — down to a single book entry on Amazon, for example — have become sophisticated in recent years. This is aided by the fact that only a few state-owned access points connect the domestic Internet to the outside world. Chinese 'netizens' counter these constraints with ingenuity, such as using Internet proxies to bypass state firewalls, or posting opinions in unrelated forums to postpone detection. The Chinese habit of reposting — in which a user copies an article in its entirety to a new forum, rather than linking to the original posting — makes the job of eradicating an erratic blog much harder.

Sixteen years ago this month, media magnate Rupert Murdoch declared that "advances in the technology of telecommunications have proved an unambiguous threat to totalitarian regimes everywhere". Last year, China overtook the United States as the country with the largest online population. In the time between, Yang's book documents how China's netizens have stumbled on online activism as a response to, among other things, a flawed justice system. Time will tell whether the revolution in communication technologies will lead to a new cultural or social revolution. ■

Li Gong is chairman and chief executive of Mozilla Online, 21 Jian Guo Men Wai Avenue, Chaoyang, Beijing 100020, China. e-mail: lgong@mozilla.com

China's unofficial democracy

The Power of the Internet in China: Citizen Activism Online

by Guobin Yang

Columbia University Press: 2009. 320 pp. \$29.50, £20.50

In July this year, a 20-year-old university student in the southern Chinese city of Hangzhou was sentenced to three years in prison for driving recklessly and killing a pedestrian. This would have been a sad but unremarkable case, except that it was only brought following a huge national outcry. Reports that local police initially protected the student, whose family was well connected, were spread over the Internet and eventually forced the police to respond.

Similar examples of online citizen activism occur every day. *The Power of the Internet in China* analyses how the Internet's rapid development in China has given its citizens a mechanism to air and share individual opinions that may differ from official positions, to connect and organize often against the will of the authorities, and to improve their own lives directly and visibly. The Internet allows Chinese citizens to practise, as cultural critic Raymond Williams termed it, "unofficial democracy".

In researching the book, Guobin Yang, a professor at Columbia University who grew up in China, read Chinese material first-hand, observed and participated in online forums and interacted with Chinese citizens online. The book's 70 case studies range from patients with diabetes or hepatitis B fighting against governmental employment discrimination, to Internet-organized worldwide demonstrations in response to the 1998 Indonesian atrocities towards the local ethnic Chinese population, to massive online and offline protests over news reporting by Western media in the run-up to the 2008 Beijing Olympics.

Yang's recounting of notable events along the historical path to China's online activism brought back old memories of my own. The first electronic gathering place targeted at people interested in China — the USENET newsgroup soc.culture.china — was started soon after I left Beijing for Cambridge, UK, in late 1987. I quickly became an active participant, devoting entire mornings to reading and replying to postings. As a student, I helped edit *China News Digest*, the first China-themed English-language electronic newsletter, which was published free by e-mail.

The milestone event for the citizens' Internet



China's online community has found its own voice.

Darwin's legacy down under

Reframing Darwin: Evolution and Art in Australia

Ian Potter Museum of Art, Melbourne, Australia

Until 1 November

Australia's unusual fauna and flora, encountered by Charles Darwin during his eagerly anticipated visit of 1836, surely influenced his evolutionary thoughts. Yet writing in his *Journal of Researches* (later known as *Voyage of the Beagle*), he focused more on Australia's human inhabitants — its convicts, settlers and Aboriginal people — than on its natural history. Nevertheless, as the exhibition *Reframing Darwin* at the Ian Potter Museum of Art in Melbourne shows, Darwin's legacy for science and art in Australia is great.

The exhibition includes diverse pieces, from fine images of HMS *Beagle* and Australia at the time of Darwin's visit, to a turn-of-the-century undergraduate exam paper containing a question about Darwinian concepts — set at the University of Melbourne by Walter Baldwin Spencer, who was appointed foundation chair of biology in 1887. Tom Roberts's powerful portrait *Aboriginal Head—Charlie Turner* (1892) conveys great emotion, which was unusual for its time, and may have been a response to Darwin's *The Expression of the Emotions in Man and Animals*. Other artworks include Emmanuel Frémiet's shocking bronze statue *Gorilla Carrying Off a Woman* — a gift in 1907 from the artist to the National Gallery of Victoria — that is juxtaposed in the gallery with Julie

Rap's unsettling digital pictures of women's bodies that have been enhanced to comment on Darwin's theory of sexual selection.

Also displayed are several works by Syms Covington, Darwin's servant aboard the *Beagle*. In a letter to his sister, Darwin offers a brief and unflattering description of Covington, who was a potential witness to the evolution of Darwin's key idea. This characterization inspired the embryo of *Mr Darwin's Shooter* (Random House, 1998), the critically acclaimed novel by Australian author Roger McDonald, which places the challenging idea of natural selection in exquisite perspective.

The precise layout of the *Beagle* is recorded in Philip Gidley King's ink sketches of the upper, lower and quarterdeck. King, an Australian-born midshipman who served aboard the vessel, drew them from memory when he was 73. His sketches were used by a Melbourne craftsman to build an exact and finely crafted replica of Darwin's modest cabin — 2.7 metres wide by 1.5 metres deep by 1.8 metres high — a workspace that Darwin shared with King, John Lort Stokes and hundreds of books.

Reframing Darwin highlights two watercolours of the *Beagle* in the imposing Chilean landscape of Valparaíso Bay. Originally attributed to artist Conrad Martens who joined the ship at Montevideo in Uruguay, inconsistencies in the palette and composition had long puzzled art historians. However, recent auctions in Santiago and London of pictures by the little-known English artist Carlos Chatworthy

Wood Taylor, also known as Charles C. Wood, suggested that the watercolours are the work of Wood, who lived in Chile for some 30 years. It remains unclear why *Beagle* captain Robert FitzRoy commissioned them.

Shown for the first time in public are ten watercolours by Louisa Anne Meredith, who was born in Birmingham, UK, in 1812 and migrated to Australia in 1839. Meredith is best known for her botanical illustrations, so the vivid images of Tasmanian fish are a surprise. The paintings' extraordinary detail challenges the view that nineteenth-century female illustrators merely pursued the picturesque. Colonial artists such as Meredith made a substantial contribution to our early understanding of Australian natural history by enhancing the lifeless specimen collections with living images. Meredith corresponded with many scientists, including the botanist

Joseph Hooker who was a friend of Darwin, and became a respected authority on Tasmanian natural history. It is remarkable that these lovely paintings have been hidden away for so long.

Colonial Tasmania's sorry history with Aboriginal people did not pass unnoticed by Darwin, who anticipated the decline in the indigenous population. Five haunting monochrome watercolours of Tasmanian Aboriginals, painted by Thomas Boch in 1837, the year after Darwin's visit, reflect a respect by the artist that contrasts with the distasteful popular views of that time. Tom Roberts's portraits of Aboriginals echo a similar sensitivity.

For those who miss this terrific exhibition, its themes are explored in a beautifully illustrated book by Jeanette Hoorn that also serves as a catalogue (*Reframing Darwin*, Miegunyah Press; 2009). It includes a biography of Baldwin Spencer, arguably Australia's first evolutionary biologist, and explains what compelled Frémiet to create his bizarre and compelling *Gorilla*. We also learn that Darwin's apparent indifference to collecting Australian flora and fauna was not down to a lack of time or interest, but to the fact that French naturalists had already done so.

Darwin's new framework for understanding life generated vigorous debate among scholars of science and letters at the time. *Reframing Darwin* stimulates that discussion once more.

Mark A. Elgar is an evolutionary biologist in the Department of Zoology, University of Melbourne, Victoria 3010, Australia
e-mail: m.elgar@unimelb.edu.au



Darwin200



Augustus Earle's 1826 painting of Australia's Blue Mountains — where Darwin walked a decade later.

For more on Darwin, see www.nature.com/darwin.



A sporadic eruption of desert art nods to humans' ability to adapt.

IAN LAUDER.COM

A creative celebration of evolution

Burning Man 2009: Evolution

Black Rock Desert, Nevada

31 August–7 September 2009

The Burning Man festival is a unique happening. For one week in September every year, the featureless Black Rock Desert in Nevada hosts a temporary community of artists, technologists and visionaries. Lacking paved roads, water, electricity and any permanent structures, Black Rock City emerges from the ephemeral lakebed, or playa, with a population of nearly 50,000. Afterwards, it disappears without trace, only to be reconfigured a year later.

Fittingly for the 2009 iteration of this social experiment, this year's theme was 'Evolution'. In the 23 years that Burning Man has been replicating, certain behaviours have been selected for by the inhabitants: radical inclusion and tolerance, self-reliance coupled with extreme altruism, a gift economy and a leave-no-trace environmental ethic. Add intense creativity, conscious participation, ingenuity and a propensity for hedonism, and the outcome is an unparalleled celebration of the human spirit.

The principal vehicle is art, from giant sculptures and lavish pyrotechnics to countless instances of the most basic art of human interaction: giving and receiving. The 'man' effigy is the centre of the festival, both figuratively and literally. This year, the 12-metre human shape hovered over a thorny forest — a tangled bank — atop a giant double helix. The DNA molecule provided a powerful artistic meme, representing both life's

capacity to evolve through genetics, and perhaps something that needs to be overcome through non-genetic evolutionary paths. Viewed from a different angle, the man seemed to float above a field of sea lilies, placing this celebration of human consciousness in an ancient evolutionary context.

The most striking image at this year's Burning Man, expressed in various ways across the city, was the famous "ascent of man" progression from great ape through to modern human, with the Burning Man icon representing the next step. This sequence resonated with the advance in human culture realized in Burning Man. One vision was the Fishbug, *Chimera sententia*, a creature rising out of the playa with an arthropod tail, amphibian body, mammalian trunk and oversized primate brain.

We created a zone at Burning Man that explored atavisms — reappearances of past events in new contexts — in human social evolution. At our Atavism Camp we created 'The Spandrel', a shade structure built with materials salvaged from the 'boneyard' at the University of Washington's Friday Harbor Marine Lab: leftover materials from past experiments, now reborn for a new purpose. At a symposium entitled 'Evolution and Society', we asked how society has interpreted evolution and whether, despite its shadowy past, its principles can guide us to a much-needed behavioural shift towards sustainability.

In the rampant transfer of culture at Burning Man, on a par with endosymbiotic events, we see hope. Evolution is evoked here on many levels: the adaptation and thriving of the



Darwin200

individual in this extreme environment, the various camps as interactive and artistic spaces, the city as it alters over the seven days and from year to year, exhibiting emergent properties of altruism, shared community and free expression. 'Burners'

become extremophiles. With resources scarce in the desert, intense sharing is the most efficient practice, suggesting that humans may yet realize a sustainable evolutionary trajectory.

Next year's theme of 'Metropolis' moves the festival a step further. Cities embody the best and worst of humanity, and Black Rock City is no exception. With its preponderance of oversized gas-guzzling camper vans, fossil-fuel-powered generators and gratuitous combustion, it is no Utopia. But the City's Alternative Energy Zone, with its huge bank of solar panels, multiple experiments in grey-water evaporation, and wind-powered cocktail bar, is paving the way.

Exodus from the barren plain brings us to the comparative paradise of juniper, sage and pinyon jays. Likewise, evolution beyond Burning Man embodies what happens off the playa, how we share and act upon our experiences. ■

Jason Hodin¹, Cory D. Bishop², Fred A. Sharpe³ and **Ruben E. Valas⁴** are evolutionary biologists.

¹Hopkins Marine Station, Stanford University, California, USA.

e-mail: seastar@stanford.edu

²Dalhousie University, Halifax, Nova Scotia,

Canada. ³Alaska Whale Foundation, Seattle, Washington, USA. ⁴University of California, San Diego, La Jolla, California, USA.

For more on evolution, see www.nature.com/darwin.

HUMAN GENETICS

Sharp focus on the variable genome

John A. L. Armour

Copy-number variation — deleted or duplicated regions of DNA — is widespread in the human genome. A systematic population survey of the common variants provides an invaluable resource for further studies.

What makes people different? Much of the answer comes from inherited differences, and interpreting the extensive variation between people's genomes is a necessary part of understanding the human genome. Variation in the form of single base changes (single nucleotide polymorphisms, SNPs), and repetitive DNA, is already well documented. Adding an extra dimension to human genetic variation is the increasingly evident prevalence and functional importance of copy-number variation. Although most human DNA is present in exactly two copies per cell — one from each parent — some regions can be variably duplicated or deleted, leading to population variation in the number of copies inherited by different individuals. In a *Nature* paper that has just appeared online, Conrad *et al.*¹ report a working map for frequent human copy-number variation. It is a landmark in providing an unprecedented combination of completeness and spatial resolution, and is likely to stand as a definitive resource for years.

This, though, is by no means the first genome-wide survey of human copy-number variation^{2–5}. Previous investigations involving a technique called array-CGH — comparative genomic hybridization to microarrays of DNA targets — have detected numerous examples of copy-number variants (CNVs)^{2,3}. Array-CGH involves hybridizing fluorescently labelled genomic DNA from the test individual simultaneously with DNA from a reference individual (also labelled, but differently) to a set of 'target' DNA sequences from different parts of the human genome. Where test and reference DNAs both have the same numbers of copies of a DNA sequence, that target will give a standard ratio of signals from the two fluorescent labels. If there is a different copy number, the ratio will shift — for example towards a lower representation of the test sample label for a region in which there is a deletion (Fig. 1).

However, because comparative hybridization has hitherto been measured using relatively large pieces of DNA, the extent of DNA involved in a deletion or duplication has often been defined imprecisely. Consequently, there are real difficulties in interpreting precise location

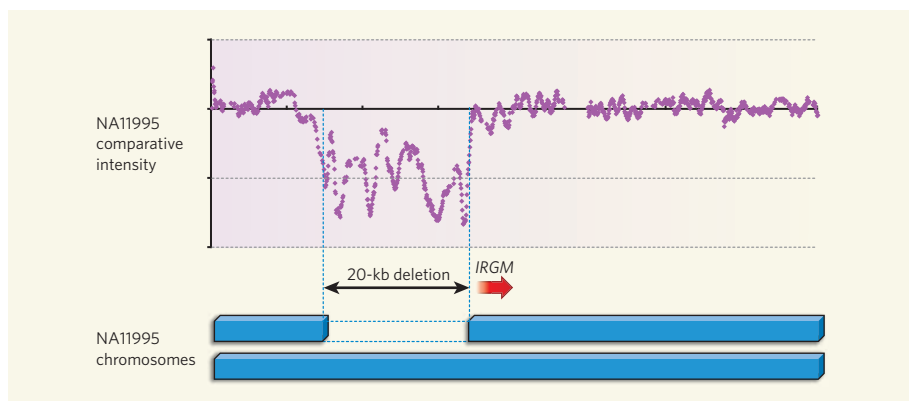


Figure 1 | Demonstrating copy-number variants (CNVs) related to disease. This CNV, revealed in Conrad and colleagues' data¹ by a sharp local reduction in comparative fluorescence intensity of a sample designated NA11995, is a deletion of about 20 kilobases that occurs upstream of the *IRGM* gene. The deletion affects one of the two copies of chromosome 5 in NA11995. It is known to influence disease predisposition, in that carriers of the deletion have a significantly increased risk of developing Crohn's disease⁷. The data for this figure were downloaded from web resources provided at www.sanger.ac.uk/cgi-bin/humgen/cnv/42mio/downloadBigDB.cgi.

in current databases of CNVs; for example, are two independent reports of CNVs in approximately the same place detecting different variants or simply rediscovering the same one? A more precise alternative for discovering CNVs uses DNA sequencing to identify non-standard sequences around the junctions of deletions or duplications^{4,5}. But even with the power of current sequencing technologies, relatively few individuals can be thoroughly surveyed using this method.

Conrad *et al.*¹ solve the problem of comprehensively defining variation at high precision by introducing a step-change in the spatial resolution of genome-wide array-CGH. Despite the problems imposed by repetitive DNA in the human genome, their survey examined comparative hybridization at no fewer than 42 million locations, using a short, synthetic DNA target for each location tested — an average spacing of about 56 base pairs. The result is comparative intensity data for each synthetic target, which can be analysed for evidence of deletion or duplication. These data were noisy (and so needed to be averaged over several neighbouring probes to be reliable), but in practice the high density of coverage closes the gap between previous hybridization approaches

and sequence-based discovery methods. This high-resolution platform was used to survey DNA from 40 unrelated individuals (20 Africans and 20 Europeans), giving a probability of better than 95% of finding CNVs present at a frequency of 5% or more.

Even applying conservative criteria for inferring CNVs, requiring ten consecutive targets to agree in reporting a deletion or duplication, nearly 12,000 putative variants were initially identified, with each individual tested differing in copy number from the reference sample at more than 1,000 distinct sites. More than 8,000 CNVs were then firmly established using a variety of validation methods — most significantly, samples from the Wellcome Trust Case-Control Consortium disease-association study⁶ were independently typed for the CNVs, the results of which will be reported separately.

Collectively, the CNVs overlap about 13% of human genes. Some deletions remove entire genes; others will cause loss of gene function via frameshifts, in which the triplet DNA coding register is shifted backwards or forwards. Deletions or duplications, especially those affecting an entire gene, have a higher a priori probability of affecting the gene's function than individual SNPs. Conrad *et al.*¹ immediately

applied their new data to investigate the potential role of CNVs in disease, by cross-checking SNPs implicated in previous human disease studies against SNPs they found to be associated with CNVs. Could a local CNV be the real cause of some of these predispositions to disease (with the SNP acting as an indirect reporter)? If so, the SNP identified as over-represented in disease should correlate with chromosomes carrying a CNV. Reassuringly, this survey for CNV–SNP–disease associations produced a list including three well-established examples — CNVs associated with Crohn's disease⁷ (Fig. 1), psoriasis⁸ and obesity⁹. Other CNVs on the list then become strong candidates for constituting the functional basis of the observed associations of SNPs with other disorders. Although these might be invaluable leads for understanding particular disorders, the authors are clear that the CNVs cannot solve the 'missing heritability' problem: in even the best-worked cases of disorders for which genetic predispositions have been characterized, most of the total risk attributable to genetic factors remains unexplained.

This study¹ has not found all human CNVs — the smallest CNVs, the less frequent CNVs and those embedded in complex, repetitive DNA will all have had a good chance of escaping detection. But Conrad *et al.*¹ will have discovered and characterized nearly all the CNVs big enough and frequent enough to matter, probably including many that will prove to be involved in disease.

The authors also provide superb resources that will allow other researchers to use their data to find out more. These include a detailed listing of the genomic locations of the CNVs found, the genotypes of reference individuals and (most useful of all) a web-based archive of (nearly) raw data from the original 40 comparative hybridization experiments. Making hybridization data freely available allows others to undertake detailed analyses of specific regions, for example to investigate potential variants not meeting the strict criteria imposed in this study. The Single Nucleotide Polymorphism database (dbSNP) and International HapMap Project provide essential data for research into SNPs. Information from this study¹ will likewise become the first-line source of CNV data for investigating human variation, genome evolution and disease genetics. ■

John A. L. Armour is at the Institute of Genetics, University of Nottingham, Queen's Medical Centre, Nottingham NG7 2UH, UK.
e-mail: john.armour@nottingham.ac.uk

1. Conrad, D. F. *et al.* *Nature* doi:10.1038/nature08516 (2009).
2. Iafrate, A. J. *et al.* *Nature Genet.* **36**, 949–951 (2004).
3. Redon, R. *et al.* *Nature* **444**, 444–454 (2006).
4. Korb, J. O. *et al.* *Science* **318**, 420–426 (2007).
5. Kidd, J. M. *et al.* *Nature* **453**, 56–64 (2008).
6. The Wellcome Trust Case Control Consortium *Nature* **447**, 661–678 (2007).
7. McCarroll, S. A. *et al.* *Nature Genet.* **40**, 1107–1112 (2008).
8. de Cid, R. *et al.* *Nature Genet.* **41**, 211–215 (2009).
9. Willer, C. J. *et al.* *Nature Genet.* **41**, 25–34 (2009).

QUANTUM MECHANICS

Passage through chaos

Daniel A. Steck

A quantum system can undergo tunnelling even without a barrier to tunnel through. The latest experiments visualize this process in exquisite detail, completely reconstructing the state of the evolving system.

Reconciling quantum mechanics with classical, Newtonian physics has been a long-standing challenge. A major aspect of this challenge pertains to chaotic systems — simple and deterministic classical systems that nevertheless display complex, seemingly random, unpredictable behaviour. The problem of 'quantum chaos' is this: take a chaotic system, study its (simplest) quantum counterpart, and what you don't find is any of the unpredictable, chaotic behaviour from the classical world. This is a funny thing, because you can go into any toy store and see any number of chaotic, pendulum-like devices, dynamically waving about for the amusement of children everywhere. In principle, a physicist should be able to model these toys either as Newtonian collections of interacting rigid bodies, or as ensembles of manifestly quantum-mechanical atoms. The answer should be the same in either case, except, however, that the chaos seems to be missing from the quantum side of the picture. Now Jessen and colleagues¹, writing on page 768 of this issue, have experimentally studied the behaviour of the quantum version of a chaotic system with an unprecedented level of precision and detail, providing new insight into the quantum–classical boundary.

To understand Jessen and colleagues' experiments¹, first consider what happens to an ensemble of atoms from the classical perspective. It is the angular momentum of the atoms that we are concerned with — technically an abstract quantity, but it suffices to think of the 'orientation' of an atom as its axis of rotation.

Fixing the magnitude of the angular momentum, we can represent the orientation of each atom as a point on a sphere (Fig. 1a). The experiments transform the orientations of the atoms in two parts: the first is a 'twist', in which a carefully tuned laser pulse shears the points on the sphere (Fig. 1b), and the second is a rotation caused by a magnetic-field pulse (Fig. 1c). The authors' sequence of twist/turn transformations on the atoms realizes for the first time the 'kicked top' (Fig. 1), one of the simplest yet most important model systems for studying quantum chaos.

The behaviour of an atom under this simple twist/turn map is rich and complex. To visualize it, consider the flattened representation of the sphere in Figure 2a, which shows the initial orientations of two groups of atoms forming two short line segments. The effect of repeating the twist/turn procedure ten times is shown in Figure 2b: one set of orientations is marginally distorted, whereas the other is stretched and folded in an intricate way. The stretching is indicative of erratic chaotic behaviour, and the point is that the dynamical behaviour in a given system can be mixed — certain initial orientations lead to chaotic dynamics, whereas others are comparatively ordered. This is best shown in Figure 2c, which plots the effects of many twist/turn iterations on several initial orientations. Chaotic regions appear as a mass of dots, whereas stable regions are neatly organized into nested, ring-like layers. The important lesson to remember for now is this: because of the stretching, an atomic orientation in the chaotic region can wander throughout

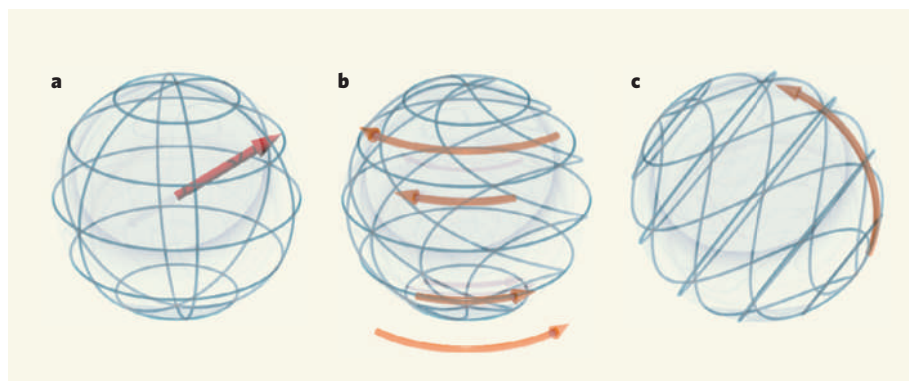


Figure 1 | The kicked top. **a**, The angular-momentum vector (arrowed) of an atom can be visualized as corresponding to a single point on a sphere, which represents all possible angular momenta. **b**, The first, or 'kick', step in realizing the 'kicked top' model system, which Jessen and colleagues¹ implement in their study of quantum chaos, is a 'twist' of the points on the sphere — the points near the poles rotating the most, and the points on the equator staying put. **c**, The second step is a simple rotation of the whole globe about an orthogonal axis (not shown).

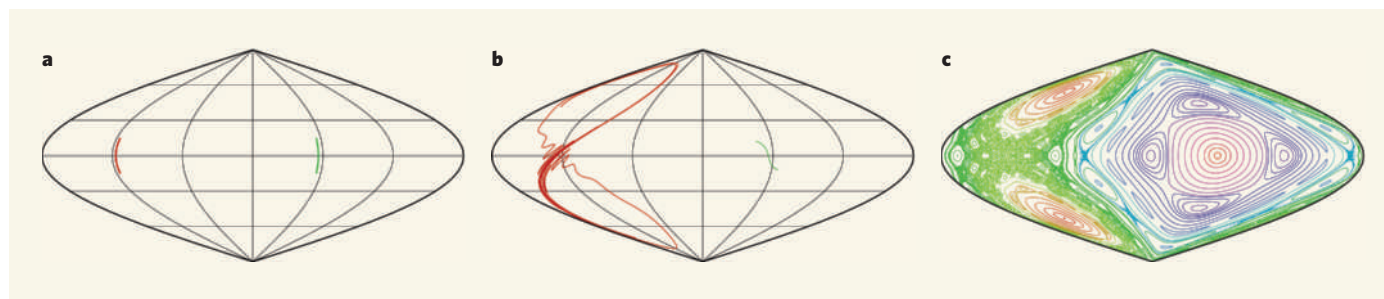


Figure 2 | Chaos in the kicked top. **a**, In this ‘flattened globe’, the two coloured line segments denote two sets of points on the sphere, each representing the initial angular momenta of two collections of atoms. **b**, The effect of ten iterations of the kicked-top transformation depicted in Figure 1: the green line segment gets only a bit twisted, whereas the red segment is dramatically stretched and folded onto itself — a hallmark of chaos. **c**, Many iterations of several starting points, clearly showing regions of stability (onion-like rings) and chaos (a fuzz of dots).

it, whereas an atom in an ‘island of stability’ is trapped there, confined to its particular ‘ring’.

But now back to quantum mechanics — we’re talking about atoms, after all. As a consequence of Heisenberg’s uncertainty principle, quantum states of atoms can’t be single points on the sphere, but must be smeared out to occupy at least some finite area. And again, there can be no chaos in the quantum case, in stark contrast to the classical model. Traditionally, there have been two approaches to this problem of the missing quantum chaos. One is to study the conditions under which the classical and quantum descriptions agree. For example, under a weak, continuous measurement, a quantum system can be persuaded to display chaos as appropriate to the classical case². The other is to study the ‘fingerprints’ of chaos³ in the quantum system, and this is the approach taken by Jessen and collaborators¹.

The authors studied a phenomenon called dynamical tunnelling⁴. This is a bit different from the better-known barrier tunnelling, in which a quantum particle can penetrate a potential barrier despite not having enough energy to hop over it. Recalling the kicked-top behaviour depicted in Figure 2c, notice that there are two main stable islands in the left hemisphere and that a consequence of stability is that, classically, an atom starting in either island is trapped there — not by any potential barrier, but merely as a consequence of the twist/turn dynamics. Because of the symmetry of these two islands, quantum mechanics allows an atom starting in one island to hop back and forth to the other island, a dynamical tunnelling process between two atomic orientations strictly forbidden in the classical world. Jessen and collaborators’ experiments clearly demonstrated this, as well as an atomic quantum state sitting placidly in the large island and another moving erratically (though not chaotically) in the chaotic region — carefully respecting the classical boundaries between stability and chaos, despite being far into the quantum regime.

The beauty of the experiments¹ lies in the complete reconstruction of the quantum state, leaving no aspect of the tunnelling process hidden. This is no easy task, involving the processing and combination of many measurements,

and was not possible in previous studies of tunnelling^{5–8}. The recovery of the full state also permitted observations of other fingerprints of chaos in a quantum system for the first time, such as the generation of quantum entanglement and the sensitivity to perturbations to the parameters of the system, rather than to its initial state⁹.

Interesting future directions for Jessen and colleagues’ work include a push towards the classical limit, where more distinct quantum states live on the sphere. This is a technically difficult regime, but one in which the fingerprints of chaos can be studied in even more detail, and where the controlled transition from quantum stability to classical chaos may be observed. ■

Daniel A. Steck is in the Oregon Center for Optics and Department of Physics, University of Oregon, Eugene, Oregon 97403-1274, USA.
e-mail: dsteck@uoregon.edu

1. Chaudhury, S., Smith, A., Anderson, B. E., Ghose, S. & Jessen, P. S. *Nature* **461**, 768–771 (2009).
2. Bhattacharya, T., Habib, S. & Jacobs, K. *Los Alamos Sci.* **27**, 110–125 (2002).
3. Berry, M. V. *Proc. R. Soc. Lond. A* **413**, 183–198 (1987).
4. Davis, M. J. & Heller, E. J. *J. Chem. Phys.* **75**, 246–254 (1981).
5. Dembowski, C. *et al. Phys. Rev. Lett.* **84**, 867–870 (2000).
6. Hensinger, W. K. *et al. Nature* **412**, 52–55 (2001).
7. Steck, D. A., Oskay, W. H. & Raizen, M. G. *Science* **293**, 274–278 (2001).
8. Steck, D. A., Oskay, W. H. & Raizen, M. G. *Phys. Rev. Lett.* **88**, 120406 (2002).
9. Peres, A. *Quantum Theory: Concepts and Methods* (Springer, 1995).

VISION

Gene therapy in colour

Robert Shapley

Replacing a missing gene in adult colour-blind monkeys restores normal colour vision. How the new photoreceptor cells produced by this therapy lead to colour vision is a fascinating question.

Colour blindness is a common genetic disorder (affecting about 5–8% of males, although fewer than 1% of females) in which the absence of a single gene on the X chromosome leads to a specific loss of function. Normal human colour vision relies on three distinct photopigments in the retina’s cone photoreceptors. Those who do not inherit the gene for one of the three cone pigments are called dichromats; such individuals cannot distinguish the difference between some pairs of colours that trichromats can discriminate easily. John Dalton, the famous British chemist, was a dichromat, and colour blindness is often referred to as daltonism.

Colour blindness is common in New World monkeys, such as the squirrel monkey (*Saimiri sciureus*), because the species does not have all three of the cone-pigment genes that humans usually have. All male and some female squirrel monkeys are colour-blind dichromats, although

most female squirrel monkeys achieve trichromatic colour vision. But let’s pay attention to squirrel monkey dichromats. Mancuso *et al.*¹ report in this issue (page 784) that injecting a virus carrying a gene for the missing photopigment into the retina of adult colour-blind squirrel monkeys confers normal trichromatic vision; 20 weeks after injection the new pigment was expressed in cone photoreceptors and the formerly dichromatic monkeys began to discriminate between two colours that had looked identical to them before treatment. Mancuso and colleagues¹ named one of their dichromatic monkeys Dalton after the chemist, but at the end of their experiments their Dalton was no longer colour blind. The success of these experiments offers promise that, perhaps in the foreseeable future, a similar therapy might improve visual function in humans. At the same time, these results raise a number of interesting questions about colour vision in primates.

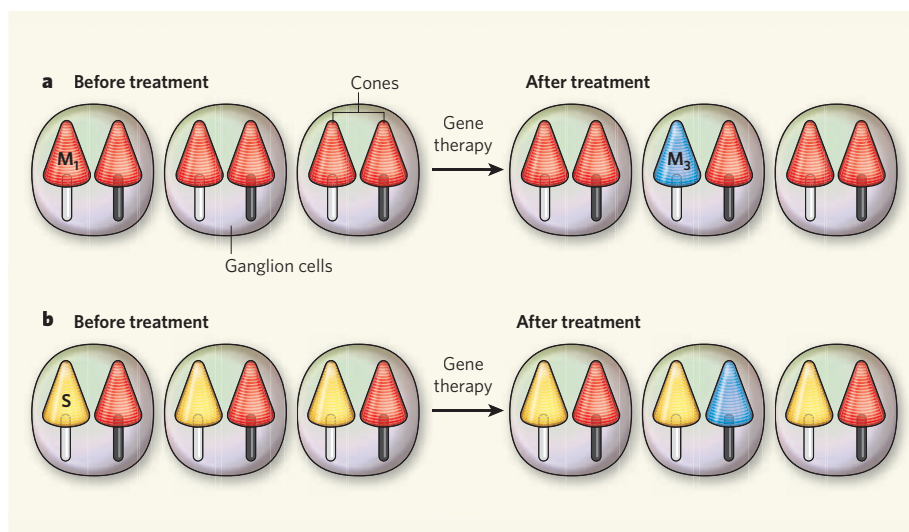


Figure 1 | Two paths to colour vision after gene therapy. The figure shows two different schemes that could generate trichromatic vision in populations of a dichromat's cone-opponent ganglion cells after gene therapy. Excitatory connections between cones and ganglion cells are indicated by white bars and inhibitory connections by black bars. **a**, The new pigment M_3 is inserted into cones that excite ganglion cells that are not cone-opponent and do not respond to colour in the dichromats (left) because they compute differences of M_1 -cone inputs only. These cells may become $M_3 - M_1$ cone-opponent cells after injection (right), producing a functional red-green pathway in treated monkeys. This is a simplified schema; a more realistic connectivity is $(M_1 + M_3) - M_1$. **b**, The new pigment M_3 can substitute for some M_1 pigment and thereby generate $S - M_3$ cells (right) that exist alongside and functionally complement the $S - M_1$ cells that are already present in the dichromat (left). Once again, a more realistic picture would show the treated cells to be $S - (M_1 + M_3)$.

In humans, monkeys and most other vertebrates, each cone photoreceptor absorbs light over a broad range of the visible spectrum and transduces it into electrical signals. We identify each cone type by its light-absorbing photopigment, which is named for the wavelength of peak absorption. Humans have short-wavelength S-cones (with peak absorption at ~440 nm), medium-wavelength M-cones (peak absorption ~535 nm) and longer-wavelength L-cones (peak absorption ~560 nm). Colour blindness in humans is usually caused by an absence of either M- or L-cones; from his symptoms, we can infer that John Dalton was missing M-cones.

Dichromatic squirrel monkeys have S-cones but only one other cone type — a middle-wavelength cone that contains only one of three possible middle-wavelength pigments (denoted M_1 , M_2 and M_3) with peak absorptions at 535, 545 and 560 nm, respectively; M_3 in the squirrel monkey is like the human L-cone pigment, M_1 like the human M-cone pigment.

There is no direct connection between the peak absorption wavelength and the role of the cone in colour perception; although L-cones (560 nm) are crucial for our ability to see red, the appearance of 560-nm light is in fact greenish-yellow. But there is no mystery about this — the signals for colour are not the signals from individual cones but rather the cone-difference signals computed by post-receptoral cells in the retina and in the brain. Signals from the cones are passed through bipolar cells to the retinal ganglion cells which lie in a deeper layer of the retina and transport visual information to

the brain. The retinal ganglion cells that carry signals about colour are called cone-opponent ganglion cells because they subtract the signals from different types of cone photoreceptor². In most mammals there are ganglion cells that subtract signals from longer-wavelength cones from the excitatory signals from S-cones, and these ganglion cells tell the difference between blue and yellow³. For example, in Old World primates the blue-yellow signal difference is usually computed as $S - (L + M)$. The blue-yellow ganglion-cell pathway in an individual squirrel monkey dichromat can be $S - M_1$, $S - M_2$ or $S - M_3$, depending on what longer-wavelength pigment the monkey has.

Humans and Old World monkeys also have red-green cone-opponent retinal ganglion cells, the responses of which are proportional to the difference between signals from L-cones and M-cones ($L - M$ or $M - L$). In trichromatic squirrel monkeys there is also a red-green pathway that computes the difference between the two longer-wavelength cones: $M_3 - M_1$, $M_3 - M_2$, or whatever pair of cones the monkey has⁴. In Old-World monkeys there are many more red-green than blue-yellow ganglion cells, but in trichromatic squirrel monkeys the blue-yellow ganglion cells are much more numerous than the red-green⁴.

Dichromatic squirrel monkeys with only M_1 -cones cannot discriminate blue-green lights with wavelengths of around 495 nm from grey light. But Mancuso and colleagues¹ report that, after therapy with the gene encoding the M_3 pigment, their treated monkeys can easily tell blue-green from grey, just like trichromats.

One of many open questions to arise from these results is: what type of cone-opponent ganglion cell is active in the treated monkeys?

There are two possibilities (Fig. 1). First, M_1 -cone signals from cones not affected by virus could be subtracted from the new M_3 -cone signals, in effect producing a new functioning red-green ($c_3 M_3 + c_1 M_1$) - M_1 pathway, where c_3 and c_1 are weighting coefficients $\neq 1$. This possibility would require that M_3 -cone signals are connected with some specificity to ganglion cells, for instance M_3 -cone signals would only be excitatory while M_1 -cone signals could remain both excitatory and inhibitory.

A second possibility is that new M_3 -cone signals could be subtracted from S-cone signals to produce a new functioning blue-yellow $S - (M_1 + M_3)$ pathway that would complement the $S - M_1$ pathway already present in the dichromat⁴. Having both $S - M_1$ and $S - (M_1 + M_3)$ cells would allow the monkey to discriminate between blue-green and grey.

One outstanding feature of Mancuso and colleagues' data¹ makes the second explanation — let's call it the blue-yellow hypothesis — more plausible. The authors monitored the time course of cone-pigment function after gene therapy by measuring cone signals in an electroretinogram (ERG), and they report that signs of new, functioning M_3 -cone pigment appeared about 20 weeks after injection. Almost simultaneously with the appearance of viable new photopigment, the formerly dichromatic monkeys became able to perform the colour-discrimination task as proficiently as trichromats. That there was no measurable delay in visual function suggests that the new cone signals were combined immediately in pre-existing colour channels from eye to brain.

The blue-yellow hypothesis would theoretically require little or no rewiring, which is why it seems more likely. But this is only speculation. The question can be answered by making electrophysiological measurements in the treated squirrel monkeys to determine whether or not there are new red-green cone-opponent retinal ganglion cells or red-green cells in the lateral geniculate nucleus^{2,4} (the first target of retinal ganglion cells), and also whether or not there are new $S - M_3$ or $S - (M_1 + M_3)$ blue-yellow cells as well as $S - M_1$ cells.

In their paper¹, Mancuso *et al.* remind us of the long-held belief that "neural connections established during development would not appropriately process an input that was not present from birth", but their results refute this idea. Their paper is a pointer to future exciting research.

Robert Shapley is at the Center for Neural Science, New York University, New York 10003, USA. e-mail: shapley@cns.nyu.edu

1. Mancuso, K. *et al.* *Nature* **461**, 784-787 (2009).
2. De Valois, R. L. *Cold Spring Harb. Symp. Quant. Biol.* **30**, 567-579 (1965).
3. Jacobs, G. H. *Phil. Trans. R. Soc. B* **364**, 2957-2967 (2009).
4. Jacobs, G. H. *Vision Res.* **23**, 461-468 (1983).

COSMOLOGY

Dark is the new black

Richard Massey

Rival experimental methods to determine the Universe's expansion are contending to become the fashionable face of cosmology. Fresh theoretical calculations make one of them the hot tip for next season.

Since the Big Bang, the Universe's initial expansion has been gradually slowed by the gravitational pull from the mass it contains. Most of this mass is in the form of invisible and mysterious dark matter. Today, however, the Universe seems to be re-accelerating under the influence of even weirder stuff dubbed dark energy. For astronomy funding purposes, 'dark' is the new black. Almost nothing is understood about either dark matter or dark energy — but both are many times more common than visible matter, and their tug of war will shape the fate of the entire cosmos.

Tracking the expansion of the Universe, from which the relative amounts of dark matter and dark energy can be inferred, requires measuring the distances to galaxies. Distances have always been the bane of astronomy: there are no simple red and green glasses to extrude our two-dimensional picture of the sky into an expanding movie. Three rival techniques are currently trying to establish themselves as the best probe of cosmological expansion. A series of calculations by Schmidt *et al.*^{1,2} now allows one contender — gravitational lensing — to predict the observational consequences of different cosmological theories at sufficient accuracy to be distinguished by future galaxy surveys.

The accelerated expansion of the Universe was first detected about a decade ago^{3,4} from observations of exploding stars called type Ia supernovae. These explosions happen at the same phase of stellar evolution, so they should all be of the same intrinsic brightness, regardless of their distance, but should look fainter the farther away they are from Earth. However, the accelerating expansion of the Universe means that distant supernovae have already receded farther from us and look even fainter. Initial enthusiasm for using supernovae as cosmic distance indicators, and thus as a probe of the Universe's expansion, garnered vast allocations of time on ground- and space-based telescopes, and triggered the first plans for a dedicated, all-sky successor to the Hubble Space Telescope. Unfortunately, the explosions were later found to depend on the stars' environment and ingredients, which evolve over cosmic time. Such effects can be parameterized only to a certain precision, and

the technique is falling out of fashion.

Distances can also be determined from the focal lengths of gravitational lenses. Gravitational lensing is the deflection of light from distant galaxies when it passes through the warped space-time around foreground 'lensing' masses along our line of sight. Just as in conventional optics, the efficiency of light deflection depends on the distance to the lens and to the source. Faraway galaxies look slightly magnified, and their shapes are dis-



Figure 1 | You thought that light travels in straight lines? Not so in the curvy world of gravitational lensing, where new results^{1,2} disentangle the zoom from the fisheye.

torted (Fig. 1). Characteristic patterns induced by lensing in the apparent shapes of distant galaxies were first observed in 2000, and were first used to constrain the properties of dark energy in 2006⁵. Measuring the subtle shape changes in distant galaxies requires a telescope with exceptional optics, which is possible only above Earth's atmosphere. But the technique was initially hailed as perfectly clean, because the only underlying physics is Einstein's well-understood theory of general relativity.

Following the same product life cycle as supernova distances, further studies revealed several potential physical flaws. First, lensing measurements assume that galaxies' true shapes are random, to infer that any observed patterns are produced entirely by light deflection around the foreground mass. However, the tidal gravitational forces between adjacent galaxies may elongate them towards each other, and one slightly in front may itself lens one slightly behind; both effects can mimic

the lensing signal and bias the measured distances. Second, lensing simultaneously distorts and enlarges galaxies, and the two effects cannot be measured independently without one biasing the other. Finally, the lensing magnification also makes a survey more likely to find highly distorted galaxies and ignore undistorted ones.

The first of the aforementioned effects, known as the intrinsic-alignments problem, can be overcome by a three-dimensional analysis of the galaxy locations, in which the alignment is measured from close pairs and then subtracted from the rest⁶. The second, called reduced shear, was solved⁷ by changing the theory to meet the data, diminishing the expected increase in the distortion signal at a given distance behind a lens that also enlarges. Schmidt *et al.*² have now performed a similar feat with the final, 'magnification bias' problem.

Schmidt and colleagues' solution² is a crucial advance for the technique of gravitational lensing, but is not without limitations. For it to work, theoretical calculations against which observations are compared must correctly predict complex statistics of the cosmic distribution of mass. Looking farther along some sight lines than others also mixes the cosmological signal with the method's built-in 'B-mode' control experiment. This had previously been used to check for potential imperfections in the telescope optics, so they now need to be even better.

Distances can also be measured by one final technique. Ripples from sound waves generated in the early Universe left their imprint on relic radiation from the Big Bang — the cosmic microwave background — and also on structures at all cosmic epochs. In patterns known as baryon acoustic oscillations, galaxies visible today are preferentially separated from each other by a set physical distance — which depends on the size of the sound-wave ripples and reliably seems to be smaller the farther away the galaxies are. This technique was not even considered worth mentioning in research proposals in the mid-1990s, but emerged in 2005 as the most important result from two large galaxy surveys — the Two-degree-Field and Sloan Digital Sky Survey⁸. Larger ground-based telescopes are currently setting out to measure this effect, but seeds of doubt are already emerging about how faithfully real galaxies trace the original ripples.

As scientific fashions come and go, the rivalry between the three houses might be more at home on the catwalks of Paris or Milan. The techniques are at different stages of the same product cycle. Initial hype draws a flurry of excitement, but when systematic physical flaws show up, sober reflection brings a sheepish look back at the design. Some methods may be consigned to a dusty

IMAGE: NASA, ESA, S. BECKWITH (STSC), HUBBLE HERITAGE TEAM (STSC/AURA)

drawer. But the stitch or two of alterations by Schmidt and colleagues^{1,2} has ensured that gravitational lensing will still be on the hot list next season.

Richard Massey is at the Institute for Astronomy, Royal Observatory Edinburgh, Blackford Hill, Edinburgh EH9 3HJ, UK. e-mail: rm@roe.ac.uk

MICROBIOLOGY

Life on leaves

Johan Leveau

The surface of plant leaves — the phyllosphere — is home to many microbes. A 'community proteogenomics' approach offers a fresh look at what it takes to survive and thrive in this unique habitat.

Under the microscope, aerial plant leaves resemble eerie landscapes, with deep gorges, tall peaks and gaping pits that riddle the waxy surface. Add to this scenery a climate that features temperature highs of 50 °C or more, exposure to harmful ultraviolet rays, erratic periods of drought and limited access to nutrients, and one gets the picture that this is a hostile environment. Still, many bacteria, fungi, yeast and other microorganisms dwell in great abundance in this 'phyllosphere'¹, which is the subject of a new investigation by Delmotte and colleagues². In their paper, published in *Proceedings of the National Academy of Sciences*, they bring twenty-first-century tools to bear on the phyllosphere, with special reference to bacteria.

Much is known about microbial adaptations to the leaf surface — for instance the production of pigments to avoid DNA damage from solar radiation or the accumulation of compatible solutes to deal with water stress. However, most of this knowledge has been inferred from single microbial species, from cultivating representative isolates in the laboratory and from exposing isolates artificially to plant foliage for an assessment of which genes contribute to microbial fitness in the phyllosphere. Delmotte and colleagues' investigation² is an exercise in 'community proteogenomics'. This approach does not rely on cultivation, does not focus on a single species, and does not suffer from the controlled conditions that typify lab experiments. The result is a snapshot-like, culture-independent insight into the diverse mechanisms that underlie the success of leaf-surface microbial colonists — in this case bacteria, the most abundant of the colonists at estimated densities of 10⁶–10⁷ cells per square centimetre (ref. 3).

Community proteogenomics⁴ arose from the marriage between metagenomics and metaproteomics. Metagenomics involves analysis of the mix of all microbial DNA present in a particular environmental sample, whereas metaproteomics does the same for all proteins.

- Schmidt, F., Rozo, E., Dodelson, S., Hui, L. & Sheldon, E. *Phys. Rev. Lett.* **103**, 1301–1304 (2009).
- Schmidt, F., Rozo, E., Dodelson, S., Hui, L. & Sheldon, E. *Astrophys. J.* **702**, 593–602 (2009).
- Perlmutter, S. *et al. Nature* **391**, 51–54 (1998).
- Riess, A. *et al. Astron. J.* **116**, 1009–1038 (1998).
- Jarvis, M. *et al. Astrophys. J.* **644**, 71–79 (2006).
- King, L. & Schneider, P. *Astron. Astrophys.* **398**, 23–30 (2003).
- Shapiro, C. *Astrophys. J.* **696**, 775–784 (2009).
- Cole, S. *et al. Mon. Not. R. Astron. Soc.* **362**, 505–534 (2005).

The metaproteomic portion of Delmotte and colleagues' approach involved collection of microbial biomass from leaf surfaces, protein extraction and digestion, separation of the fragments by liquid chromatography and analysis by mass spectrometry. The result was a mixed bag of nearly half a million spectra, each corresponding to a short peptide sequence. Linking these spectra to proteins with a possible function and evolutionary origin is a challenge and is possible only with a proper frame of reference. Typically, this frame is provided by the publicly available databases of annotated DNA and protein sequences.

However, if a microbial community has few representatives in the public database, the chances are that many of the sequences in the database will be too dissimilar to allow positive matching with short peptide sequences from the environmental proteome. This is where the metagenomic part of the proteogenomic approach comes in: it increases the probability of protein identification by metagenomic profiling of the same sample from which the proteins were extracted. In the case of Delmotte *et al.*², pyrosequencing was used to construct a representative library of DNA sequences from the leaf samples: by including these metagenomic data on top of the sequences in the public database, up to 87% more proteins could be identified in the bacterial leaf communities. This suggests that many bacteria from the leaves of the plants that were investigated — soya bean, clover and *Arabidopsis* — are indeed genetically distinct from the bacteria for which genomic data are currently available. This was especially true for members of the genus *Sphingomonas*, which were among the most numerous bacteria present. Were it not for the metagenomic data, none of the abundant proteins assigned to this genus would have been identified.

The phyllosphere metaproteome reveals that many of the highly expressed bacterial proteins — porins, TonB-like proteins and components of ABC-type transporters,

for example — are apparently involved in scavenging what little food there is available on the leaf surface. This possibility is consistent with studies showing the limited access to nutrients in the phyllosphere, such as the products of photosynthesis that leak from the leaf interior⁵. Proteins for using methanol, a plant waste product, were also abundant and could be assigned to *Methylobacterium* species — leaf colonizers of many different plants⁶. Stress proteins were over-represented as well, revealing a need to protect the bacterial cells from oxidative and osmotic damage, and to prevent them from becoming desiccated. One of the surprising finds was the prominence of a protein containing a fasciclin domain, possibly involved in cell adhesion, but with no previously suspected role in survival in the phyllosphere.

The wider context for this line of research is illustrated by considering the significance of microbial populations on leaves. They play a part in the global nitrogen and carbon cycles; they participate in removing airborne pollutants⁷; and they contribute to the decomposition of leaf litter and to the production of plant and animal feed by composting and silaging. The traditional focus of phyllosphere research has been on microorganisms that are of agricultural relevance, in particular plant pathogens and their antagonists. But the discovery of archetypal leaf bacteria such as the plant-pathogenic *Pseudomonas syringae* in non-agricultural environments⁸, and the detection of human enteropathogens such as *Escherichia coli* O157:H7 on leaf surfaces⁹, are inviting a more expansive view of the phyllosphere as a source and sink of environmental bacteria.

One value of the study by Delmotte *et al.*² is that it will help to draw this microbial habitat to the attention of a broader audience of researchers and into the field of comparative 'omics'. It will also serve as a baseline for further proteogenomic excursions into the phyllosphere, which are likely to involve studies at higher resolution, both temporally and spatially. Among the issues to be addressed are the dynamics of microbial protein expression relative to changes in community composition, and the role of the plant and its environment in driving the functional plasticity of foliage-associated microorganisms.

Johan Leveau is in the Department of Plant Pathology, University of California, Davis, California 95616-8751, USA. e-mail: jleveau@ucdavis.edu

- Ruinen, J. *Plant Soil* **15**, 81–109 (1961).
- Delmotte, N. *et al. Proc. Natl Acad. Sci. USA* **106**, 16428–16433 (2009).
- Lindow, S. E. & Brandl, M. T. *Appl. Environ. Microbiol.* **69**, 1875–1883 (2003).
- VerBerkmoes, N. C., Denef, V. J., Hettich, R. L. & Banfield, J. F. *Nature Rev. Microbiol.* **7**, 196–205 (2009).
- Leveau, J. H. J. & Lindow, S. E. *Proc. Natl Acad. Sci. USA* **98**, 3446–3453 (2001).
- Sy, A. *et al. Appl. Environ. Microbiol.* **71**, 7245–7252 (2005).
- Sandhu, A., Halverson, L. J. & Beattie, G. A. *Environ. Microbiol.* **9**, 383–392 (2007).
- Morris, C. E. *et al. Infect. Genet. Evol.* **7**, 84–92 (2007).
- Brandl, M. T. *Annu. Rev. Phytopathol.* **44**, 367–392 (2006).

STRUCTURAL BIOLOGY

Tracing Argonaute binding

Samir Bouasker and Martin J. Simard

Argonaute proteins inhibit gene expression by binding to messenger RNA via a small nucleic-acid guide. Structures of the Argonaute complex bound to target RNA reveal snapshots of a silencing machine at work.

Argonaute proteins are essential regulators of gene expression, being key components of gene-silencing pathways mediated by RNA. In plants and animals, Argonaute proteins interact with small RNA molecules to form an enzymatically active complex called RISC. The RISC complex silences target genes by binding to messenger RNA that has sequence complementarity to the Argonaute-bound RNA. Some Argonaute proteins cleave mRNA; others have lost their catalytic activity and regulate gene expression by some other means, most likely by inhibiting mRNA translation¹.

Although previous biochemical studies have revealed much about the molecular mechanisms of RNA silencing, exactly how Argonaute proteins interact with small RNA molecules as well as with their target mRNAs has been unclear. The paper by Wang and colleagues² from the Patel and Tuschl labs (page 754 of this issue) is therefore especially welcome — the authors report the results of a series of impressive structural studies that reveal the molecular dynamics of the Argonaute protein as it binds to, and slices up, its RNA target.

Argonaute proteins have four domains: the amino-terminal, PAZ, MID and PIWI domains (Fig. 1). The PAZ domain is implicated in the binding of single-stranded RNA, whereas the PIWI domain has endonucleolytic activity — the ability to cleave nucleic acid at internal bonds. The first crystal structures of Argonaute proteins from prokaryotes^{3–5} (bacteria or archaea) revealed that the PIWI domain is structurally similar to members of the RNase H

family of endonucleases, which use DNA as a guiding template to target RNA molecules. Although prokaryotic Argonaute proteins use DNA as the nucleic-acid template strand, plant and animal Argonaute proteins have evolved to use single-stranded RNA, rather than DNA, as a template to target RNA.

The Patel and Tuschl labs recently reported⁶ the structure of a binary complex of bacterial Argonaute bound to a 5'-phosphorylated DNA guide strand. In this work, they showed that both ends of the 21-nucleotide DNA guide strand are anchored within the Argonaute protein — the 5'-phosphate end is anchored in a binding pocket in the MID domain, and two nucleotides at the extreme 3'-hydroxyl end of the guide DNA strand are anchored in a pocket of the PAZ domain. The same two groups described⁷ the first ternary complex of bacterial Argonaute, consisting of the Argonaute protein bound to guide DNA plus a 20-nucleotide target RNA. To solve this ternary structure, Patel, Tuschl and colleagues had to prevent Argonaute from cleaving the target RNA; they did this by introducing base mismatches between the guide DNA and the target RNA. However, this manoeuvre made it impossible to assess guide DNA–target RNA base-pairing at and beyond the mispaired cleavage site. So although this study⁷ showed that the Argonaute complex undergoes conformational changes when it binds to target RNA, the researchers could not clearly observe the behaviour of Argonaute during the nucleation and propagation of the guide DNA–target RNA

complex or during cleavage of target RNA.

In their latest study, the authors² circumvent this problem by using a catalytic mutant of bacterial Argonaute, which is unable to slice up target RNA. The authors analysed several structures of the mutant Argonaute protein bound to a guide DNA that is fully base-paired to target RNA molecules of different lengths. As previously observed, both ends of the guide strand are anchored in the Argonaute binding pockets (Fig. 1a).

The nucleation step begins with the binding of the guide DNA to target RNA at the DNA 5'-phosphate end (Fig. 1b). The guide DNA and target RNA form base pairs and 'zipper up', forming a DNA–RNA double helix that extends from the 5' end of the guide DNA to DNA nucleotide 16. Pivoting of the Argonaute protein allows double-helix formation while both ends of the guide DNA are anchored in the Argonaute binding sites. Beyond nucleotide 16, Argonaute's N-terminal domain blocks additional base-pairing towards the 3'-hydroxyl end of the guide strand. The propagation of base-pairing between the guide DNA and the target RNA before this obstruction results in the release of the 3' end of the DNA from its anchor site in the PAZ domain (Fig. 1c). This release allows rotation of the PAZ domain, leading to a conformational change that favours the positioning of the cleavage site of target RNA close to the catalytic residues in the PIWI domain (Fig. 1d).

The structures solved by Wang *et al.*² also clearly show that two magnesium ions that are essential for cleavage activity are located one on either side of the cleavage site. This requirement for cations to facilitate site-specific cleavage is also a feature of RNase H endonucleases, confirming that Argonaute cleavage activity is highly similar to that of RNase H enzymes.

Although studies of Argonaute proteins in prokaryotes are informative, prokaryotic Argonaute has not been implicated in small RNA-mediated silencing pathways. The next big challenge will be to solve the structures of

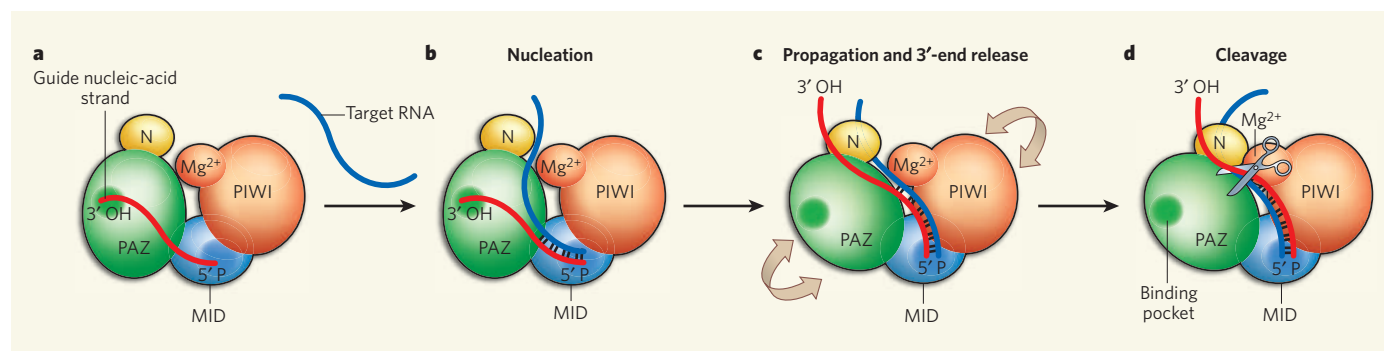


Figure 1 | The Argonaute silencing complex at work. **a**, Argonaute proteins have four domains: the amino-terminal domain (N), PAZ, MID and PIWI. Each Argonaute protein binds to a small nucleic-acid molecule (red; RNA in plants and animals, and DNA in bacteria), which functions as a template for binding to complementary target RNA. The 5'-phosphate (5' P) end of the guide nucleic acid is anchored in the MID domain, and the 3'-hydroxyl end (3' OH) is anchored in the PAZ domain. **b**, Structural studies by Wang and colleagues² reveal that when the Argonaute complex binds to target RNA, the nucleation step begins with formation of a double helix by base pairing between the guide nucleic acid and the target RNA, commencing at the 5'-phosphate end of the guide strand. **c**, Pivotal movement of the Argonaute protein allows extension of the double helix while the guide DNA is anchored at both ends. The 3'-hydroxyl end of the guide strand is then released from the PAZ domain, allowing its rotation. **d**, This conformational change favours the exact positioning of the target RNA cleavage site close to the Argonaute PIWI domain. Magnesium ions in the PIWI domain facilitate precise cleavage of the target.

Argonaute proteins in plants and animals. It will be interesting to know whether plant and animal Argonaute proteins promote nucleation of all of the guide strand's nucleotides with the RNA target to increase binding and silencing specificity, or whether they nucleate only up to position 16, like bacterial Argonaute. Another question is whether interaction between animal microRNA (a type of small RNA encoded in the genome that is used as a guide strand) and target mRNA can be accommodated by the Argonaute protein, because microRNA typically binds imprecisely to target mRNA and forms an imperfectly paired RNA double helix. Comparing the structural features of Argonaute proteins from different organisms will help us to further

understand their functions within the RNA silencing pathways and might even uncover new roles for this versatile protein family. ■
Samir Bouasker and Martin J. Simard are at the Laval University Cancer Research Centre, Hôtel-Dieu de Québec, Quebec City, Quebec G1R 2J6, Canada.
e-mail: martin.simard@crhdq.ulaval.ca

1. Hutvagner, G. & Simard, M. J. *Nature Rev. Mol. Cell Biol.* **9**, 22–32 (2008).
2. Wang, Y. *et al. Nature* **461**, 754–761 (2009).
3. Song, J.-J., Smith, S. K., Hannon, G. J. & Joshua-Tor, L. *Science* **305**, 1434–1437 (2004).
4. Yuan, Y.-R. *et al. Mol. Cell* **19**, 405–419 (2005).
5. Ma, J.-B. *et al. Nature* **434**, 666–670 (2005).
6. Wang, Y. *et al. Nature* **456**, 209–213 (2008).
7. Wang, Y. *et al. Nature* **456**, 921–926 (2008).

PHOTONICS

One-way road for light

Eli Yablonovitch

The transmission of information from one place to another by light waves sent through waveguides is hampered by light attenuation and scattering loss. Magnetic photonic crystals could provide a solution to such problems.

The concept of photonic crystals — periodically arranged structures specifically engineered to trap and guide light — grew from an initial analogy^{1–3} with the electronic band structure of semiconductors. In these materials, no electrons can be found that have energies within a range known as the ‘band gap’. Similarly, in photonic crystals, photons whose frequencies fall within the ‘photonic band gap’ are prevented from flowing inside the material.

Haldane and Raghu⁴ have recently extended the equivalence between the behaviour of photons in photonic crystals and that of electrons in electronic systems. They have predicted the photonic analogue of the ‘edge’ states that characterize the quantum Hall effect⁵ that is experienced by the electrons of a two-dimensional (2D) electron gas when it is subjected to a strong magnetic field. Under certain conditions, photons can be confined to the edges of a 2D photonic crystal — one whose lattice structure has 2D periodicity — and be restricted to unidirectional propagation. On page 772 of this issue, Wang *et al.*⁶ report observing such photonic edge states in a magneto-optical 2D photonic crystal, verifying Haldane and Raghu's theoretical prediction⁴.

To achieve unidirectional photonic edge states requires a system that lacks time-reversal symmetry — that is, one with physical properties that are not preserved by a time-reversal transformation. To realize such a ‘non-reciprocal’ system, Wang and colleagues⁶ used a photonic crystal consisting of a 2D-periodic arrangement of magneto-optical ferrite rods; the magneto-optical nature of the rods

confers the desired time-reversal asymmetry on the system. After characterizing the system's band gap, the authors demonstrated the unidirectional character of the system's edge states: forward-propagating transmission outweighed backward-propagating transmission by almost 50 decibels.

Wang and colleagues' experimental demonstration⁶ of the correspondence between the optics of a photonic crystal and the elegant physics of the quantum Hall effect is not only a delight for fundamental science, it also opens the door to practical applications based on non-reciprocal photonic crystals. These crystals may provide the means to develop a new type of optical-fibre waveguide that would be utterly immune to energy loss caused by scattering from material defects or obstacles.

Photonic-crystal fibres⁷, a form of optical fibre based on 2D photonic crystals, have been very successful in providing unique functions⁸ in fibre-optic communications. The most interesting type of photonic-crystal fibre has a hollow core in which light is confined by a surrounding cladding that consists of either a 2D-periodic photonic crystal or concentric ‘Bragg rings’⁹. Because their cores are hollow rather than being filled with a material substance, light channelling through them suffers less absorption loss, enabling low-loss propagation over long distances. Indeed, it has been shown¹⁰ that photonic-crystal fibres can achieve very low loss. But they are not quite as lossless as one would hope owing to scattering caused by the intrinsic roughness of their internal (glass) cladding surfaces¹⁰: the lowest

reported loss is still about a factor of ten larger than that of their best conventional counterparts. In transoceanic optical-fibre systems, underwater amplifiers must be placed approximately every 100 kilometres to compensate for loss.

One of the distinctive properties of 3D-periodic photonic crystals is that light is confined in all directions. As a consequence, and unlike in ordinary fibres or 2D-periodic photonic-crystal fibres, light travelling through hollow waveguides carved out of 3D-periodic photonic crystals is not subject to scattering loss. This increases the possibility of attaining ultra-low-loss light propagation, with both absorption and scattering losses suppressed. However, it does not prevent back scattering. Light can propagate both forwards and backwards within the same hollow waveguide, and back scattering off an obstacle within the waveguide can reduce forward transmission, and so be a source of loss even in a 3D-periodic photonic crystal.

Haldane and Raghu's theoretical ideas⁴, together with the experiments of Wang *et al.*⁶, now offer a solution to the back-scattering problem. By using a magneto-optical, photonic-crystal system that breaks time-reversal symmetry, Wang and colleagues show that it is possible to design the material's dispersion relationship, which describes the way in which wave propagation varies with frequency, such that, for a given frequency band, only forward-propagating waves exist. The ferrites the authors⁶ used operate at microwave, rather than optical, frequencies. Nonetheless, there are several other magneto-optical materials that are used in the optical regime.

To sum up, the ideal optical waveguide would be made of a low-loss hollow core, with a layer of non-reciprocal material, surrounded by a 3D-periodic photonic crystal, providing immunity to both back scattering and surface-roughness scattering. With such a low-loss waveguide, the possibility would then exist for one-hop transoceanic communication across 10,000 kilometres — about the distance from San Francisco to Tokyo — without the current requirement for electronic repeaters or amplifiers. ■

Eli Yablonovitch is in the Electrical Engineering and Computer Sciences Department, University of California, Berkeley, California 94720-1770, USA.
e-mail: eliy@eecs.berkeley.edu

1. Yablonovitch, E. J. *Opt. Soc. Am. B* **10**, 283–295 (1993).
2. John, S. *Phys. Rev. Lett.* **58**, 2486–2489 (1987).
3. Yablonovitch, E. *Phys. Rev. Lett.* **58**, 2059–2062 (1987).
4. Haldane, F. D. M. & Raghu, S. *Phys. Rev. Lett.* **100**, 013904 (2008).
5. von Klitzing, K., Dorda, G. & Pepper, M. *Phys. Rev. Lett.* **45**, 494–497 (1980).
6. Wang, Z., Chong, Y., Joannopoulos, J. D. & Soljačić, M. *Nature* **461**, 772–775 (2009).
7. Russell, P. St J. *J. Lightwave Tech.* **24**, 4729–4749 (2006).
8. www.nktpotonics.com
9. Yeh, P., Yariv, A. & Marom, E. *J. Opt. Soc. Am.* **68**, 1196–1201 (1978).
10. Roberts, P. *et al. Opt. Express* **13**, 236–244 (2005).

Finding the missing heritability of complex diseases

Teri A. Manolio¹, Francis S. Collins², Nancy J. Cox³, David B. Goldstein⁴, Lucia A. Hindorf⁵, David J. Hunter⁶, Mark I. McCarthy⁷, Erin M. Ramos⁵, Lon R. Cardon⁸, Aravinda Chakravarti⁹, Judy H. Cho¹⁰, Alan E. Guttmacher¹, Augustine Kong¹¹, Leonid Kruglyak¹², Elaine Mardis¹³, Charles N. Rotimi¹⁴, Montgomery Slatkin¹⁵, David Valle⁹, Alice S. Whittemore¹⁶, Michael Boehnke¹⁷, Andrew G. Clark¹⁸, Evan E. Eichler¹⁹, Greg Gibson²⁰, Jonathan L. Haines²¹, Trudy F. C. Mackay²², Steven A. McCarroll²³ & Peter M. Visscher²⁴

Genome-wide association studies have identified hundreds of genetic variants associated with complex human diseases and traits, and have provided valuable insights into their genetic architecture. Most variants identified so far confer relatively small increments in risk, and explain only a small proportion of familial clustering, leading many to question how the remaining, 'missing' heritability can be explained. Here we examine potential sources of missing heritability and propose research strategies, including and extending beyond current genome-wide association approaches, to illuminate the genetics of complex diseases and enhance its potential to enable effective disease prevention or treatment.

Many common human diseases and traits are known to cluster in families and are believed to be influenced by several genetic and environmental factors, but until recently the identification of genetic variants contributing to these 'complex diseases' has been slow and arduous¹. Genome-wide association studies (GWAS), in which several hundred thousand to more than a million single nucleotide polymorphisms (SNPs) are assayed in thousands of individuals, represent a powerful new tool for investigating the genetic architecture of complex diseases^{1,2}. In the past few years, these studies have identified hundreds of genetic variants associated with such conditions and have provided valuable insights into the complexities of their genetic architecture^{3,4}.

The genome-wide association (GWA) method represents an important advance compared to 'candidate gene' studies, in which sample sizes are generally smaller and the variants assayed are limited to a selected few, often on the basis of imperfect understanding of biological pathways and often yielding associations that are difficult to replicate^{5,6}. GWAS are also an important step beyond family-based linkage studies, in which inheritance patterns are related to several hundreds to thousands of genomic markers. Despite many clear successes in single-gene 'Mendelian' disorders^{7,8}, the limited success of linkage studies in complex diseases has been attributed to their low power and resolution for variants of modest effect^{9–11}.

The underlying rationale for GWAS is the 'common disease, common variant' hypothesis, positing that common diseases are attributable in part to allelic variants present in more than 1–5% of the population^{12–14}. They have been facilitated by the development of commercial 'SNP chips' or arrays that capture most, although not all, common variation in the genome. Although the allelic architecture of some conditions, notably age-related macular degeneration, for the most part reflects the contributions of several variants of large effect (defined loosely here as those increasing disease risk by twofold or more), most common variants individually or in combination confer relatively small increments in risk (1.1–1.5-fold) and explain only a small proportion of heritability—the portion of phenotypic variance in a population attributable to additive genetic factors³. For example, at least 40 loci have been associated with human height, a classic complex trait with an estimated heritability of about 80%, yet they explain only about 5% of phenotypic variance despite studies of tens of thousands of people¹⁵. Although disease-associated variants occur more frequently in protein-coding regions than expected from their representation on genotyping arrays, in which over-representation of common and functional variants may introduce analytical biases, the vast majority (>80%) of associated variants fall outside coding regions, emphasizing the importance of including both coding and non-coding regions in the search for disease-associated variants³.

¹National Human Genome Research Institute, Building 31, Room 4B09, 31 Center Drive, MSC 2152, Bethesda, Maryland 20892-2152, USA. ²National Institutes of Health, Building 1, Room 126, MSC 0148, Bethesda, Maryland 20892-0148, USA. ³Departments of Medicine and Human Genetics, University of Chicago, Room A612, MC 6091, 5841 South Maryland Avenue, Chicago, Illinois 60637, USA. ⁴Duke University, The Institute for Genome Sciences and Policy (IGSP), Box 91009, Durham, North Carolina 27708, USA. ⁵National Human Genome Research Institute, Office of Population Genomics, Suite 4076, MSC 9305, 5635 Fishers Lane, Rockville, Maryland 20892-9305, USA. ⁶Department of Epidemiology, Harvard School of Public Health, 677 Huntington Avenue, Boston, Massachusetts 02115, USA. ⁷University of Oxford, Oxford Centre for Diabetes, Endocrinology and Metabolism, Churchill Hospital, Old Road, Oxford OX3 7LJ, UK, and Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK. ⁸GlaxoSmithKline, 709 Swedeland Road, King of Prussia, Pennsylvania 19406, USA. ⁹McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, 733 North Broadway BRB579, Baltimore, Maryland 21205, USA. ¹⁰Yale University, Department of Medicine, Division of Digestive Diseases, 333 Cedar Street, New Haven, Connecticut 06520-8019, USA. ¹¹deCODE Genetics, Sturlugata 8, Reykjavik IS-101, Iceland. ¹²Lewis-Sigler Institute for Integrative Genomics, Howard Hughes Medical Institute, and Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey 08544, USA. ¹³The Genome Center, Washington University School of Medicine, 4444 Forest Park Avenue, Campus Box 8501, Saint Louis, Missouri 63108, USA. ¹⁴National Human Genome Research Institute, Center for Research on Genomics and Global Health, Building 12A, Room 4047, 12 South Drive, MSC 5635, Bethesda, Maryland 20892-5635, USA. ¹⁵Department of Integrative Biology, University of California, 3060 Valley Life Science Building, Berkeley, California 94720-3140, USA. ¹⁶Stanford University, Health Research and Policy, Redwood Building, Room T204, 259 Campus Drive, Stanford, California 94305, USA. ¹⁷Department of Biostatistics, University of Michigan, 1420 Washington Heights, Ann Arbor, Michigan 48109-2029, USA. ¹⁸Department of Molecular Biology and Genetics, 107 Biotechnology Building, Cornell University, Ithaca, New York 14853, USA. ¹⁹Howard Hughes Medical Institute and University of Washington, Department of Genome Sciences, 1705 North-East Pacific Street, Foege Building, Box 355065, Seattle, Washington 98195-5065, USA. ²⁰University of Queensland, School of Biological Sciences, Goddard Building, Saint Lucia Campus, Brisbane, Queensland 4072, Australia. ²¹Vanderbilt University, Center for Human Genetics Research, 519 Light Hall, Nashville, Tennessee 37232-0700, USA. ²²Department of Genetics, North Carolina State University, Box 7614, Raleigh, North Carolina 27695, USA. ²³Department of Genetics, Harvard Medical School, 77 Avenue Louis Pasteur, NRB 0330, Boston, Massachusetts 02115, USA. ²⁴Queensland Institute of Medical Research, 300 Herston Road, Brisbane, Queensland 4006, Australia.

The questions arise as to why so much of the heritability is apparently unexplained by initial GWA findings, and why it is important. It is important because a substantial proportion of individual differences in disease susceptibility is known to be due to genetic factors, and understanding this genetic variation may contribute to better prevention, diagnosis and treatment of disease. It is important to recognize, however, that few investigators expected these studies immediately to find all of the variants associated with common diseases, or even most of them; the hope was that they would at least find some¹⁶. Limitations in the design of early GWAS, such as imprecise phenotyping and the use of control groups of questionable comparability, may have reduced estimates of effect sizes while preserving some ability to identify associated variants¹⁷. These studies have considerably surpassed early expectations, reproducibly identifying hundreds of variants in many dozens of traits, but for many traits they have explained only a small proportion of estimated heritability¹⁸.

Many explanations for this missing heritability have been suggested, including much larger numbers of variants of smaller effect yet to be found; rarer variants (possibly with larger effects) that are poorly detected by available genotyping arrays that focus on variants present in 5% or more of the population; structural variants poorly captured by existing arrays; low power to detect gene–gene interactions; and inadequate accounting for shared environment among relatives. Consensus is lacking, however, on approaches and priorities for research to examine what has been termed ‘dark matter’ of genome-wide association—dark matter in the sense that one is sure it exists, can detect its influence, but simply cannot ‘see’ it (yet). Here we examine potential sources of missing heritability and propose research strategies to illuminate the genetics of complex diseases.

Heritability and allelic architecture of complex traits

It is reasonable to assume that allelic architecture (number, type, effect size and frequency of susceptibility variants) may differ across traits, and that missing heritability may take a different form for different diseases¹⁹, but at present our understanding is too limited to distinguish these possibilities. Age-related macular degeneration may provide the best example of a common disease in which heritability is substantially explained by a small number of common variants of large effect²⁰, but for other conditions, such as Crohn’s disease, the proportion of heritability explained is not nearly so large despite a much larger number of identified variants²¹ (Table 1). There are no obvious differences between these two traits in genetic architecture as predicted from clinical and epidemiological data that would explain the differences observed in their allelic architecture. Some apparent differences may simply be due to differences in the stage of investigation across traits. Studies in several conditions have clearly demonstrated that the number of detected variants increases with increasing sample size^{22–24}.

Population genetic theory suggests an explanation for the paucity of variants explaining a large proportion of disease predisposition, in that decreased reproductive fitness should typically act to reduce the frequencies of high-risk variants. This might explain the relative lack of variants detected so far for some neuropsychiatric conditions, such as autism spectrum disorders, given their low reproductive fitness²⁵. Yet for a condition such as type 1 diabetes, which has a similar prevalence, familial risk, early onset and poor reproductive fitness (at

least before the discovery of insulin therapy), more than 40 loci have already been reported; this might be because the overall sample sizes studied in type 1 diabetes have been very large^{26,27}. Present-day reproductive fitness may correlate poorly with the forces that have shaped variation throughout human evolution; moreover focusing on the reproductive effects of a single disease ignores the pleiotropic effects (effects of the same variant on multiple characteristics or disease risks) of multiple alleles influencing that condition simultaneously with many other conditions²⁸.

Selection might also be responsible for keeping genetic effect sizes low, as variants of larger effect may be selected against and eventually disappear¹⁹. Long-term stabilizing selection minimizes the production of individuals at the extremes of a trait²⁹, in part by reducing the additive genetic effects of alleles already present or those arising *de novo* by mutation³⁰ to levels potentially beneath the ability of studies of feasible size to detect them. Selection may also contribute to differences in the ability to detect loci in different complex diseases, if genetic susceptibility to some diseases is more strongly affected by selection than other diseases, or if environmental perturbations vary in intensity across diseases. Immune and infectious agents have been recognized as among the strongest selection pressures in human evolution³¹, and immune-related genes have been strongly implicated in Crohn’s disease and other immune-mediated diseases³, suggesting either that pleiotropic effects of these variants reduce the efficiency of negative selection, or that strong environmental perturbation in modern societies might expose the disease risk associated with these variants. Selection may thus explain why disease allele frequencies are low and allelic effects are small, but this should manifest as low, rather than missing, heritability.

A probable contributor to the small genetic effect sizes observed so far is that current investigations have incompletely surveyed the potential causal variants within each gene. Relative risks observed for marker SNPs may underestimate the actual risks associated with the true causal variants. Notably, 11 out of 30 genes implicated as carrying common variants associated with lipid levels also carry known rare alleles of large effect identified in Mendelian dyslipidemias, including *ABCA1*, *PCSK9* and *LDLR*^{22,32}, suggesting that genes containing common variants with modest effects on complex traits may also contain rare variants with larger effects.

An important consideration is that the overwhelming majority of GWAS and other genetic studies have been limited to European ancestry populations, whereas genetic variation is greatest in populations of recent African ancestry², and studies in non-Europeans have yielded intriguing new variants^{33,34}. Studies of populations of recent African ancestry in particular is likely to increase the yield of rare variants and narrow the large chromosomal regions of association identified in the ‘younger’ population due to extended linkage disequilibrium, or the tendency for adjacent genetic loci to be inherited together³¹. Isolated populations may also be of value given their potential to be enriched in unique variants³⁵.

The accuracy of current heritability estimates is also important, because experimentally identified variants could never explain all the variance in an erroneously inflated heritability estimate. Heritability of quantitative traits, formally defined as the proportion of phenotypic variance in a population attributable to additive genetic factors (narrow-sense heritability, h^2 (ref. 36)) is typically estimated from

Table 1 | Estimates of heritability and number of loci for several complex traits

Disease	Number of loci	Proportion of heritability explained	Heritability measure
Age-related macular degeneration ⁷²	5	50%	Sibling recurrence risk
Crohn’s disease ²¹	32	20%	Genetic risk (liability)
Systemic lupus erythematosus ⁷³	6	15%	Sibling recurrence risk
Type 2 diabetes ⁷⁴	18	6%	Sibling recurrence risk
HDL cholesterol ⁷⁵	7	5.2%	Residual* phenotypic variance
Height ¹⁵	40	5%	Phenotypic variance
Early onset myocardial infarction ⁷⁶	9	2.8%	Phenotypic variance
Fasting glucose ⁷⁷	4	1.5%	Phenotypic variance

* Residual is after adjustment for age, gender, diabetes.

family studies, and can be expected to vary across environments. Narrow-sense heritability estimates in humans can be inflated if family resemblance is influenced by non-additive genetic effects (dominance and epistasis, or gene–gene interaction), shared familial environments, and by correlations or interactions among genotypes and environment^{36,37}. However, heritabilities estimated from pedigree studies in animals agree well with heritability estimated from response to artificial selection, suggesting that estimates from family studies are not necessarily inflated.

Teasing apart the contributions to heritability of environmental factors shared among relatives will soon be possible because the availability of genome-wide markers now provides empirical estimates of identity-by-descent (IBD) allele sharing between pairs of relatives. For example, full sibs share on average half their genetic complement, but this proportion can vary—in one large study it ranged from 0.37 to 0.62 (ref. 38). By relating phenotypic differences to the observed IBD sharing fraction among sib pairs, marker data were used to generate a heritability estimate of 0.8 for height³⁸. This is remarkably consistent with estimates using traditional methods but free of their assumptions, suggesting that for height at least, heritability is not overestimated. Applying such estimation to distantly related or ‘unrelated’ individuals is now feasible using dense genomic scans³⁹; given the number of people with dense genotyping data, heritability estimates could be generated for a wide variety of traits free of potential confounding by unmeasured shared environment.

Improving estimates of all contributors to heritability will facilitate determination of the proportion of genetic variance that has been explained. Despite imprecision in current estimates, it may still be possible to know that ‘all the heritability’ has been explained by predicting phenotypes in a new set of individuals from trait-associated markers, and correlating the predicted phenotypes with the actual values. If the markers truly explain all the additive genetic variance, the squared correlation between predicted and actual phenotype will be equal to the heritability⁴⁰. Population-based heritability estimates thus provide a valuable metric for completeness of available genetic risk information, but individualized disease prevention and treatment will ultimately require identifying the variants accounting for risk in a given individual rather than on a population basis.

Rare variants and unexplained heritability

Much of the speculation about missing heritability from GWAS has focused on the possible contribution of variants of low minor allele frequency (MAF), defined here as roughly $0.5\% < \text{MAF} < 5\%$, or of rare variants ($\text{MAF} < 0.5\%$). Such variants are not sufficiently frequent to be captured by current GWA genotyping arrays^{14,41}, nor do they carry sufficiently large effect sizes to be detected by classical linkage analysis in family studies (Fig. 1). Once MAF falls below 0.5%, detection of associations becomes unlikely unless effect sizes are very large, as in monogenic conditions. For modest effect sizes, association testing may require composite tests of overall ‘mutational load’, comparing frequencies of mutations of potentially similar functional effect in cases and controls.

Low frequency variants could have substantial effect sizes (increasing disease risk two- to threefold) without demonstrating clear Mendelian segregation, and could contribute substantially to missing heritability⁴². For example, 20 variants with risk allele frequency of 1% and allelic odds ratio (or probability of an event occurring divided by the probability of it not occurring, compared in people with versus without the risk allele) of three would account for most familial aggregation of type 2 diabetes. There are relatively few examples of such variants contributing to complex traits, possibly owing to insufficiently large sample sizes or insufficiently comprehensive arrays.

The primary technology for the detection of rare SNPs is sequencing, which may target regions of interest, or may examine the whole genome. ‘Next-generation’ sequencing technologies, which process millions of sequence reads in parallel, provide monumental increases in speed and volume of generated data free of the cloning biases and

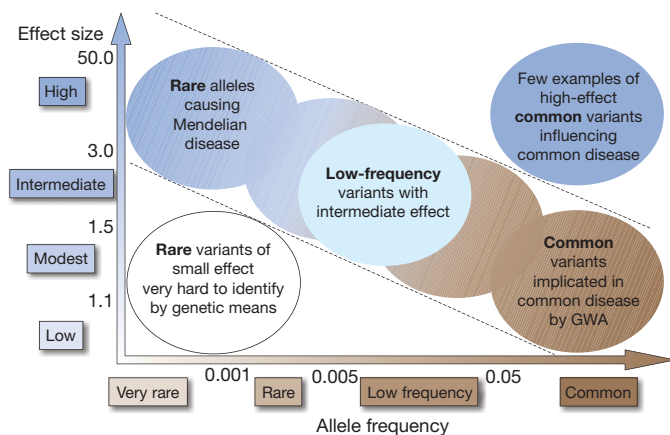


Figure 1 | Feasibility of identifying genetic variants by risk allele frequency and strength of genetic effect (odds ratio). Most emphasis and interest lies in identifying associations with characteristics shown within diagonal dotted lines. Adapted from ref. 42.

arduous sample preparation characteristic of capillary sequencing⁴³. Detection of associations with low frequency and rare variants will be facilitated by the comprehensive catalogue of variants with $\text{MAF} \geq 1\%$ being generated by the 1,000 Genomes Project (<http://www.1000genomes.org/page.php>), which will also identify many variants at lower allele frequencies. The pilot effort of that program has already identified more than 11 million new SNPs in initially low-depth coverage of 172 individuals⁴⁴.

Current mechanisms for using sequencing to identify rare variants underlying or co-located with GWA-defined associations include sequencing in genomic regions defined by strong and repeatedly replicated associations with common variants, and sequencing a larger fraction of the genome in people with extreme phenotypes. In the absence of GWA-defined signals, sequencing candidate genes in subjects at the extremes of a quantitative trait (such as lipid levels or the age at onset), can identify other associated variants, both common and rare^{45,46}. An important finding from these studies is that much of the information is provided by people at the extremes of trait distributions, who seem to be more likely to carry loss-of-function alleles⁴⁷.

Sample sizes used for the initial identification of DNA sequence variants have generally been modest, and sample size requirements increase essentially linearly with $1/\text{MAF}$. Much larger samples are needed for the identification of associations with variants than those needed for the detection of the variants themselves. They also scale roughly linearly with $1/\text{MAF}$ given a fixed odds ratio and fixed degree of linkage disequilibrium with genotyped markers. Sample size for association detection also scales approximately quadratically with $1/(\text{OR} - 1)$, and thus increases sharply as the odds ratio (OR) declines. Sample size is even more strongly affected by small odds ratios than by small MAF, so low frequency and rare variants will need to have higher odds ratios to be detected.

Complicating matters further, numerous rare variants may be detected in a gene or region but they may have disparate effects on phenotype. Common variants have typically been analysed individually^{23,48}, but with one or two carriers of each rare variant, pooling them using specific criteria becomes attractive^{47,49,50}. Pooling variants of similar class increases the effective MAF of the class and reduces the number of tests performed, but raises several other questions (Box 1).

Determining which of the multitude of variants carried by an individual are responsible for a given phenotype represents a massive task, especially if the causal alleles are relatively anonymous in terms of known functional consequences. Because only a small proportion will have obvious functional consequences for the resultant protein, lesser evidence of association may suffice to implicate variants of this sort. The best approaches for combining functional credibility and statistical support in the evaluation of such variants remain to be

Box 1 | Research strategies using rare and low frequency variants and structural variants

Research strategies using rare and low frequency and structural variants include: (1) using expanding catalogues of human sequence variation⁴⁴, by linkage disequilibrium of rare/low frequency/structural variants with GWA-genotyped SNPs and/or improved detection methods, to identify variants underlying association signals identified by SNP arrays. (2) Improving approaches for using common SNPs to predict and control for differences in rare and low frequency SNPs. (3) Using targeted sequencing judiciously, focusing on people with extreme or unusual phenotypes. (4) Including populations of recent African ancestry in sequencing studies to increase yield of rare variants and narrow large linkage disequilibrium blocks; consider isolated or founder populations potentially enriched with unique variants. (5) Focusing discovery efforts on well-phenotyped groups, accessible families with large sibships, and families that allow return to family members for iterative phenotyping. (6) Increasing emphasis on other structural variants such as inversions and translocations. (7) Implementing chromosomal-region-specific matching throughout the genome, to select for each case and for each part of their genome—a control that is more similar to the case within that genomic region rather than matching genome-wide using measures such as geographic ancestry. (8) Pooling rare variants for analysis using logical criteria, by addressing the questions: do the different rare variants increase or decrease disease risk? What classes of variants should be pooled? What is the optimal level of MAF for pooling? (9) Improving CNV detection by developing more extensive population databases in large cohorts to understand allele and mutation frequency, inheritance among unaffected individuals, and CNV calling algorithms.

determined. GWAS have tended to focus almost exclusively on statistical evidence and de-emphasize considerations of biological plausibility, but the challenges of sifting through the millions of rare variants in which two individuals differ may prompt a return to biology if rare variants are to be grouped and analysed properly.

The sheer number of inter-individual differences, mostly rare, to be detected by whole-genome sequencing (roughly 0.4% of 3 billion base pairs⁵¹) also raises the question of finding appropriate comparison subjects, or allelic matches, because people carrying rare variants at some loci may have important differences in ancestry or other factors from a general population. To reduce the number of variants that must be considered in a case-control comparison it would be useful to implement chromosomal-region-specific matching throughout the genome, to select closely related alleles and regions from the comparison population, thereby greatly reducing the number of incidental allelic differences from cases.

Structural variation and unexplained heritability

Structural variation, including copy number variants (CNVs, such as insertions and deletions) and copy neutral variation (such as inversions and translocations), may account for some of the unexplained

heritability if those variants contribute to the genetic basis of human disease and are incompletely assessed by commercial SNP genotyping arrays. Although this type of variation has not been explicitly examined in most GWAS until now, CNVs in particular (regions 1 kilobase (kb) or longer present in variable numbers across individuals) have gained attention as methods to detect them have improved^{52,53}. Other forms of structural variation such as inversions, translocations, microsatellite repeat expansions, insertions of new sequence, and complex rearrangements have been implicated in rare Mendelian conditions. For the most part such variation has been largely unexplored in relation to complex traits⁵⁴.

Variation due to CNVs arises from a combination of rare and common alleles; as with SNPs most variants are rare but most of the differences between any two individuals arise from a limited set of common (MAF ≥ 5%) copy number polymorphisms (CNPs)⁵⁵. Disease-associated CNVs detected so far, like disease-associated SNPs, include rare variants with large associated effect sizes, and common variants with more modest effects but carried by a large proportion of the population (Table 2). An added twist is that rare, highly penetrant CNVs have generally been large (600 kb–3 megabases (Mb), affecting many genes), whereas disease-associated common CNPs have been much smaller (20–45 kb) and have identified specific genomic features for follow-up study. Because both rare and common CNVs are under-ascertained by current methods, the relative affect of these variants will continue to be an important research question for CNVs just as for SNPs. Of note, CNVs arising *de novo* in current cases and shown to be of importance in neuropsychiatric and developmental conditions^{56–58} will not contribute to family resemblance and heritability, but could explain some of the variation at present attributed to ‘environment’.

Several approaches have been developed for integrating analysis of CNVs into GWAS, including innovation in the design of GWA arrays (with associated discoveries in neuropsychiatric disorders^{59,60}) and the use of the linkage disequilibrium relationships between SNPs and common CNPs (with associated discoveries in Crohn’s disease and body weight^{52,61}). These approaches are early in their development and have important limitations, although rapid progress is expected as CNV detection algorithms evolve and large-scale sequencing studies produce comprehensive, high-resolution maps of segregating CNPs that can be measured in large reference panels.

Many GWA data sets already have sufficient genotype and intensity information to permit calling of large, rare CNVs even if specific CNV probes were not included. As with non-structural single nucleotide sequence variants, more detailed (‘iterative’) phenotyping in relatives may reveal subtle phenotypic effects that were not initially appreciated.

Harnessing family studies

Family studies provide several opportunities for the investigation and interpretation of as-yet-unidentified genetic variation of many types

Table 2 | Selected disease associations with rare CNVs and common CNPs

Disease	Locus	Type of CNV	Size (kb)	Population frequency	Case frequency	Effect size (OR)
Rare CNVs						
Autism/IMR ⁵⁹	16p11.2	<i>De novo</i> deletion	600	1 × 10 ^{−4}	1%	100
Autism ⁵⁹	16p11.2	Rare duplication	600	3 × 10 ^{−4}	0.50%	16
Schizophrenia ^{60,78}	1q21.1	Rare deletion	1,400	2 × 10 ^{−4}	0.30%	15
IMR ⁷⁹	1q21.1	Rare deletion	1,400	2 × 10 ^{−4}	0.47%	Not observed in 4,737 controls
Schizophrenia ^{60,78}	15q13.3	Rare deletion	1,600	2 × 10 ^{−4}	0.20%	12
Epilepsy ⁸⁰	15q13.3	Rare deletion	1,600	2 × 10 ^{−4}	1.0%	Not observed in 3,699 controls
IMR ^{79,81}	15q13.3	Rare deletion	1,600	2 × 10 ^{−4}	0.30%	Not observed in 960 controls
Schizophrenia ⁸²	22q11.2	Rare deletion	3,000	2.5 × 10 ^{−4}	1%	40
Common CNPs						
Crohn’s disease ⁸³	<i>IRGM</i>	Deletion polymorphism	20	7%	11%	1.5
Body mass index ⁶¹	<i>NEGR1</i>	Deletion polymorphism	45	65%	Quantitative trait	<1 kg
Psoriasis ⁸⁴	<i>LCE3C</i>	Deletion polymorphism	30	55%	65%	1.3

IMR, idiopathic mental retardation.

underlying complex diseases (Box 2). Family studies may facilitate the detection of rare and low frequency variants, and the identification of their associations with common diseases, because predisposing variants will be present at much higher frequency in affected relatives of an index case.

Family studies also permit the investigation of parent-of-origin-specific effects, as have been reported for structural variants^{62,63}. If not properly accounted for, such effects could mask associations and diminish the proportion of heritability explained. High-density SNP data in extended pedigrees can be used to localize predisposition genes, as unexpectedly long runs of identity-by-state sharing among affected relatives suggest true IBD that is probably due to an underlying genetic cause⁶⁴. Linkage data can also enhance the power of high-density GWA scans by essentially relaxing *P*-value thresholds in the few instances in which suggestive findings overlap but are not definitive⁶⁵. Family studies may also be useful in identifying gene–gene interactions, because affected relatives are more likely to share two nearby epistatic loci in linkage disequilibrium that would be unlinked in unrelated individuals^{66,67}.

Strategies for existing and future GWAS

The nearly 400 GWAS published so far represent a wealth of data on the genetics of complex diseases⁴. These studies have provided valuable insights into the genetics of common diseases, particularly about the underlying genetic architecture of complex traits and the predominance of non-coding variants that may have a role in their aetiology. Just as linkage studies demonstrated that complex diseases cannot be explained by a small number of rare variants with large effects, GWAS have shown that they cannot be explained by a limited number of common variants of moderate effect (Fig. 1). The distinction between low frequency and truly rare alleles is largely an operational one, relating to the potential, given realistic effect sizes, for detecting associations with low frequency variants by GWAS at attainable sample sizes. Low frequency variants of intermediate effect might also contribute to explaining missing heritability that should be tractable through large meta-analyses and/or imputation of genome-wide association data.

GWAS will probably remain an efficient way of investigating the remaining heritability, because their association signals may well define the genomic regions where rare variants, structural variants, and other forms of underlying variation are likely to cluster. The value of future studies can be enhanced by expanding to non-European samples and less common diseases and including more precise phenotypes and measures of environmental exposures^{48,68} (Box 3). Information on lower frequency alleles emerging from projects such as the 1,000 Genomes will be used to produce even more comprehensive GWA arrays, and will facilitate the investigation of the lower frequency spectrum without the need for *de novo* sequencing.

Box 2 | Using family studies to investigate missing heritability

To investigate missing heritability using family studies, the following measures are required: (1) examine phenotypic effects of rare variants, particularly for subtle phenotypic abnormalities. (2) Investigate mutation rates and inheritance patterns of recurrent mutations. (3) Assess inheritance patterns of rare and structural variants. (4) Investigate parent-of-origin-specific effects. (5) Enhance power for identifying associated loci by studying affected sibs, particularly for conditions with substantial genetic heterogeneity. (6) Identify associated loci by unexpectedly long runs of identity-by-state sharing among distantly related affected relatives. (7) Enhance power of GWA scans by up-weighting *P* values in preselected regions based on linkage signals. (8) Identify gene–gene interactions by positive correlations between family-specific logs odds ratio (lod) scores or evidence of linkage disequilibrium among unlinked loci.

Potential of research to explain missing heritability

GWAS were initially designed to focus on the higher end of the frequency–effect size spectrum, so much work remains to be done, both in finding other variants in the lower frequency and larger effect domains shown in Fig. 1, and in understanding their functional and pathophysiological properties. To the extent that there are several causal variants on a common haplotype or that causal variants are in imperfect linkage disequilibrium with genotyped markers, marker SNPs will underestimate the associated disease risk.

The modest size of genetic effects detected so far confirms the multifactorial aetiology of these conditions and suggests that complex diseases will require substantially greater research effort to detect additional genetic influences. Near-term approaches for finding missing heritability on which there seems to be wide agreement include: targeted or whole-genome sequencing in people with extreme phenotypes, especially those with available family members and consent for recontact and iterative phenotyping; use of expanded reference panels of genomic variation such as 1,000 Genomes to enhance coverage of existing and future GWAS; mining of existing GWAS for associations with structural variants and evidence of gene–gene interactions; improved methods for detection of CNVs and other structural variants, applied to large, well-phenotyped groups and families; and expansion of sample sizes for numerous complex diseases through larger individual studies and meta-analyses, including people of non-European ancestry.

Given all that has been learned of the genetic architecture of common diseases in the past few years, it may also be worthwhile to attempt exhaustive characterization of some well-studied traits by cataloguing all the contributing variation, be it in DNA sequence, DNA structure, chromatin structure, environmental modifiers, and defining all its functional implications. Potential criteria for deciding which traits to pursue aggressively in this way might include the strength and robustness of detected associations, evidence that associations are disrupted by varying linkage disequilibrium patterns, documented associations of identified loci with multiple traits, and public health importance of the traits to be studied.

Box 3 | Making the most of existing and future GWAS

The following steps can be used to make the most of existing and future GWAS: (1) ensure the wide availability of data with appropriate protections for consent and privacy. (2) Increase sample sizes and ensure thorough meta- and mega-analyses of comparable data, with increased focus on conditions with relatively small sample sizes studied so far. (3) Expand studies to non-European samples and more diverse diseases. (4) Improve phenotyping by expanding to subtler or more quantitative or precise phenotypes as needed to reduce heterogeneity or explore pleiotropic effects. (5) Capture larger proportion of variation in implicated genes. (6) Enhance the investigation of the X chromosome, particularly as the methods for imputation of X and Y markers improve. (7) Investigate gene–gene interactions, including dominance and epistasis. (8) Investigate gene–environment interactions: measure environment rigorously and analyse it against GWA data; examine rare exposures in common diseases for unusual responders; consider including GWA in monozygotic twins or migrant studies to identify gene–environment interaction interactions; conduct suitably large (several hundred thousand people) prospective cohort studies with GWA genotyping, and reproducible reliable exposure measures at baseline; include routine biobanking of material suitable for epigenetic analysis, such as non-immortalized lymphocytes for DNA methylation or cryopreserved cell or nuclear preparations for chromatin studies; relate quantitative phenotypes to epigenetic variation, which unlike SNPs is inherently quantitative; measure epigenetic variants in appropriate tissues when technically feasible. (9) Measure CNVs: use linkage disequilibrium patterns of SNP data and improved maps and imputation methods to identify common CNPs; use SNP intensity data to identify large CNVs where feasible regions; use best possible CNP typing array until using next generation sequencing for this purpose becomes feasible.

Explaining missing heritability, however intellectually satisfying, will probably have fewer practical applications as an end in itself than as a means to an end. The ultimate goal of this line of research, as with nearly all research in the genetics of complex disease, is to improve understanding of human physiology and disease aetiology so that more effective means of diagnosis, treatment and prevention can be developed. If a genetic variant(s) was found that opened the door to effective new treatments at low cost and with minimal side effects (LDL-receptor mutations and the statin class of drugs comes to mind), one would probably be content to leave some heritability unexplained. It is the expectation that associations identified by GWAS or other genomic methods will eventually enable effective disease prevention or treatment, either through delineation of the functional properties of variants recognized at present, or identification of new variants in which true functionality lies, that primarily motivates the hunt for missing heritability.

It is more difficult to imagine predictive variants accounting for a sizeable proportion of disease risk without also explaining a sizeable proportion of heritability, and the limited incremental value in disease prediction of variants identified so far suggests that genetic prediction of complex diseases on a population basis will be challenging^{69–71}. Still, the identification of even many hundreds of risk variants of small effect should permit identification of the small proportion of a population at the highest genetically defined risk, in which targeted prevention strategies should be explored. If testing of such variants was to be conducted across several diseases, as is now feasible with dense genome-wide association genotyping and will be greatly facilitated by whole-genome sequencing, a sizeable number of people could be identified to be at greatly increased risk for at least one disease. Identification of genetic variants that influence disease risk, prognosis, or the response to treatment should enable the development of diagnostic and interventional strategies that are safe, effective and as necessary, individualized⁷¹, although the value of genetic variants in disease prediction and the steps needed to realize this are widely debated^{69,70}. Given how little has actually been explained of the demonstrable genetic influences on most common diseases, despite identification of hundreds of associated genetic variants, the search for missing heritability provides a potentially valuable path towards further discoveries.

- Hardy, J. & Singleton, A. Genomewide association studies and human disease. *N. Engl. J. Med.* **360**, 1759–1768 (2009).
- International HapMap Consortium. A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851–861 (2007).
- Hindorf, L. A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl Acad. Sci.* **106**, 9362–9367 (2009).
- Comprehensive analysis of genomic annotations for disease-associated SNPs defined by GWAS, showing great majority of associated loci in intronic or intergenic regions of unknown function.
- Hindorf, L. A., Junkins, H. A., Mehta, J. P. & Manolio, T. A. A catalog of published genome-wide association studies. Available at <http://www.genome.gov/26525384> (accessed, 18 September 2009).
- Hirschhorn, J. N., Lohmueller, K., Byrne, E. & Hirschhorn, K. A comprehensive review of genetic association studies. *Genet. Med.* **4**, 45–61 (2002).
- Todd, J. A. Statistical false positive or true disease pathway? *Nature Genet.* **38**, 731–733 (2006).
- Corder, E. H. *et al.* Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* **261**, 921–923 (1993).
- Lifton, R. P. Genetic dissection of human blood pressure variation: common pathways from rare phenotypes. *Harvey Lect.* **100**, 71–101 (2004).
- Altshuler, J., Palmer, L. J., Fischer, G., Scherh, H. & Wjst, M. Genome-wide scans of complex human diseases: true linkage is hard to find. *Am. J. Hum. Genet.* **69**, 936–950 (2001).
- Risch, N. & Merikangas, K. The future of genetic studies of complex human diseases. *Science* **273**, 1516–1517 (1996).
- Risch, N. J. Searching for genetic determinants in the new millennium. *Nature* **405**, 847–856 (2000).
- Reich, D. E. & Lander, E. S. On the allelic spectrum of human disease. *Trends Genet.* **17**, 502–510 (2001).
- Collins, F. S., Guyer, M. S. & Chakravarti, A. Variations on a theme: cataloging human DNA sequence variation. *Science* **278**, 1580–1581 (1997).
- Pritchard, J. K. Are rare variants responsible for susceptibility to common diseases? *Am. J. Hum. Genet.* **69**, 124–137 (2001).
- Visscher, P. M. Sizing up human height variation. *Nature Genet.* **40**, 489–490 (2008).
- Collins, F. S. 2005 William Allan Award address. No longer just looking under the lamppost. *Am. J. Hum. Genet.* **79**, 421–426 (2006).
- Pearson, T. A. & Manolio, T. A. How to interpret a genome-wide association study. *J. Am. Med. Assoc.* **299**, 1335–1344 (2008).
- Maher, B. Personal genomes: The case of the missing heritability. *Nature* **456**, 18–21 (2008).
- Pritchard, J. K. & Cox, N. J. The allelic architecture of human disease genes: common disease-common variant or not? *Hum. Mol. Genet.* **11**, 2417–2423 (2002).
- Jakobsdottir, J., Gorin, M. B., Conley, Y. P., Ferrell, R. E. & Weeks, D. E. Interpretation of genetic association studies: markers with replicated highly significant odds ratios may be poor classifiers. *PLoS Genet.* **5**, e1000337 (2009).
- Barrett, J. C. *et al.* Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nature Genet.* **40**, 955–962 (2008).
- Kathiresan, S. *et al.* Common variants at 30 loci contribute to polygenic dyslipidemia. *Nature Genet.* **41**, 56–65 (2009).
- Zeggini, E. *et al.* Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nature Genet.* **40**, 638–645 (2008).
- Ahmed S. *et al.* Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2. *Nature Genet.* **41**, 585–590 (2009).
- Lord, C., Cook, E. H., Leventhal, B. L. & Amaral, D. G. Autism spectrum disorders. *Neuron* **28**, 355–363 (2000).
- Cooper, J. D. *et al.* Meta-analysis of genome-wide association study data identifies additional type 1 diabetes risk loci. *Nature Genet.* **40**, 1399–1401 (2008).
- Barrett, J. C. *et al.* Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nature Genet.* **41**, 703–707 (2009).
- Keller, M. C. & Miller, G. Resolving the paradox of common, harmful, heritable mental disorders: which evolutionary genetic models work best? *Behav. Brain Sci.* **29**, 385–404 (2006).
- Gibson, G. & Wagner, G. Canalization in evolutionary genetics: a stabilizing theory? *Bioessays* **22**, 372–380 (2000).
- Gibson, G. Decanalization and the origin of complex disease. *Nature Rev. Genet.* **10**, 134–140 (2009).
- Campbell, M. C. & Tishkoff, S. A. African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. *Annu. Rev. Genomics Hum. Genet.* **9**, 403–433 (2008).
- Lusis, A. J. & Pajukanta, P. A treasure trove for lipoprotein biology. *Nature Genet.* **40**, 129–130 (2008).
- Zheng, W. *et al.* Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. *Nature Genet.* **41**, 324–328 (2009).
- Yasuda, K. *et al.* Variants in KCNQ1 are associated with susceptibility to type 2 diabetes mellitus. *Nature Genet.* **40**, 1092–1097 (2008).
- Sabatti, C. *et al.* Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. *Nature Genet.* **41**, 35–46 (2009).
- Falconer, D. S. & Mackay, T. F. C. *Introduction to Quantitative Genetics* Addison 123 (Wesley Longman Ltd, 1996).
- Visscher, P. M., Hill, W. G. & Wray, N. R. Heritability in the genomics era—concepts and misconceptions. *Nature Rev. Genet.* **9**, 255–266 (2008).
- Detailed review of strengths, weaknesses and controversies in estimations of heritability from human, agricultural and experimental studies.
- Visscher, P. M. *et al.* Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. *PLoS Genet.* **2**, e41 (2006).
- Meuwissen, T. H., Hayes, B. J. & Goddard, M. E. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157**, 1819–1829 (2001).
- Lee, S. H., van der Werf, J. H., Hayes, B. J., Goddard, M. E. & Visscher, P. M. Predicting unobserved phenotypes for complex traits from whole-genome SNP data. *PLoS Genet.* **4**, e1000231 (2008).
- McCarthy, M. I. & Hirschhorn, J. N. Genome-wide association studies: potential next steps on a genetic journey. *Hum. Mol. Genet.* **17** (R2), R156–R165 (2008).
- Insightful review of initial findings from GWAS, the heritability that they do and do not explain, and potential for progress from other GWAS, identification of rare variants, and studies of epigenetics and gene expression and function.
- McCarthy, M. I. *et al.* Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nature Rev. Genet.* **9**, 356–369 (2008).
- Mardis, E. R. The impact of next-generation sequencing technology on genetics. *Trends Genet.* **24**, 133–141 (2008).
- Abecasis, G. R. The 1000 Genomes Project: analysis of pilot datasets. *Biology of Genomes* page 246 (Cold Spring Harbor Laboratory, 5–9 May 2009).
- Kotowski, I. K. *et al.* A spectrum of PCSK9 alleles contributes to plasma levels of low-density lipoprotein cholesterol. *Am. J. Hum. Genet.* **78**, 410–422 (2006).
- Cohen, J. C. *et al.* Multiple rare variants in NPC1L1 associated with reduced sterol absorption and plasma low-density lipoprotein levels. *Proc. Natl Acad. Sci. USA* **103**, 1810–1815 (2006).
- Romeo, S. *et al.* Population-based resequencing of ANGPTL4 uncovers variations that reduce triglycerides and increase HDL. *Nature Genet.* **39**, 513–516 (2007).
- Haiman, C. A. *et al.* Multiple regions within 8q24 independently affect risk for prostate cancer. *Nature Genet.* **39**, 638–644 (2007).

49. Nejentsev, S., Walker, N., Riches, D., Egholm, M. & Todd, J. A. Rare variants of *IFIH1*, a gene implicated in antiviral responses, protect against type 1 diabetes. *Science* **324**, 387–389 (2009).
Four rare variants in *IFIH1* independently lowering risk of type 1 diabetes were identified by sequencing exons and splice sites of 10 genes under GWA-defined peaks, demonstrating the power of intensive sequencing to identify potentially causative variants in follow-up of GWAS.
50. Li, B. & Leal, S. M. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am. J. Hum. Genet.* **83**, 311–321 (2008).
51. Crawford, M. H. *Anthropological Genetics: Theory, Methods and Applications* 341 (Cambridge Univ. Press, 2006).
52. McCarroll, S. A. Extending genome-wide association studies to copy-number variation. *Hum. Mol. Genet.* **17** (R2), R135–R142 (2008).
53. Scherer, S. W. *et al.* Challenges and standards in integrating surveys of structural variation. *Nature Genet.* **39** (suppl.), S7–S15 (2007).
54. Kidd, J. M. *et al.* Mapping and sequencing of structural variation from eight human genomes. *Nature* **453**, 56–64 (2008).
55. McCarroll, S. A. *et al.* Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nature Genet.* **40**, 1166–1174 (2008).
Initial map of CNVs demonstrating high proportion (>80%) of inter-individual differences in copy number differences due to common CNVs of MAF 5% or greater; >99% of CNVs probably derived from inheritance rather than *de novo* mutation; and most common diallelic CNVs in strong linkage disequilibrium with common SNPs.
56. de Vries, B. B. *et al.* Diagnostic genome profiling in mental retardation. *Am. J. Hum. Genet.* **77**, 606–616 (2005).
57. Sebat, J. *et al.* Strong association of *de novo* copy number mutations with autism. *Science* **316**, 445–449 (2007).
58. Xu, B. *et al.* Strong association of *de novo* copy number mutations with sporadic schizophrenia. *Nature Genet.* **40**, 880–885 (2008).
59. Weiss, L. A. *et al.* Association between microdeletion and microduplication at 16p11.2 and autism. *N. Engl. J. Med.* **358**, 667–675 (2008).
60. Stefansson, H. *et al.* Large recurrent microdeletions associated with schizophrenia. *Nature* **455**, 232–236 (2008).
61. Willer, C. J. *et al.* Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nature Genet.* **41**, 25–34 (2009).
62. Abrahams, B. S. & Geschwind, D. H. Advances in autism genetics: on the threshold of a new neurobiology. *Nature Rev. Genet.* **9**, 341–355 (2008).
63. Kong, A. *et al.* Detection of sharing by descent, long-range phasing and haplotype imputation. *Nature Genet.* **40**, 1068–1075 (2008).
64. Thomas, A., Camp, N. J., Farnham, J. M., Allen-Brady, K. & Cannon-Albright, L. A. Shared genomic segment analysis. Mapping disease predisposition genes in extended pedigrees using SNP genotype assays. *Ann. Hum. Genet.* **72**, 279–287 (2008).
65. Roeder, K., Bacanu, S. A., Wasserman, L. & Devlin, B. Using linkage genome scans to improve power of association in genome scans. *Am. J. Hum. Genet.* **78**, 243–252 (2006).
66. MacLean, C. J., Sham, P. C. & Kendler, K. S. Joint linkage of multiple loci for a complex disorder. *Am. J. Hum. Genet.* **53**, 353–366 (1993).
67. Zhao, J., Jin, L. & Xiong, M. Test for interaction between two unlinked loci. *Am. J. Hum. Genet.* **79**, 831–845 (2006).
68. Waters, K. M. *et al.* Generalizability of associations from prostate cancer genome-wide association studies in multiple populations. *Cancer Epidemiol. Biomarkers Prev.* **18**, 1285–1289 (2009).
69. Clayton, D. G. Prediction and interaction in complex disease genetics: experience in type 1 diabetes. *PLoS Genet.* **5**, e1000540 (2009).
70. Khoury, M. J. *et al.* The scientific foundation for personal genomics: recommendations from a National Institutes of Health-Centers for Disease Control and Prevention multidisciplinary workshop. *Genet. Med.* **11**, 559–567 (2009).
71. Pharoah, P. D., Antoniou, A. C., Easton, D. F. & Ponder, B. A. Polygenes, risk prediction, and targeted prevention of breast cancer. *N. Engl. J. Med.* **358**, 2796–2803 (2008).
72. Maller, J. *et al.* Common variation in three genes, including a noncoding variant in CFH, strongly influences risk of age-related macular degeneration. *Nature Genet.* **38**, 1055–1059 (2006).
73. International Consortium for Systemic Lupus Erythematosus Genetics (SLEGEN). Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in *ITGAM*, *PXK*, *KIAA1542* and other loci. *Nature Genet.* **40**, 204–210 (2008).
74. Zeggini, E. *et al.* Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nature Genet.* **40**, 638–645 (2008).
75. Kathiresan, S. *et al.* Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nature Genet.* **40**, 189–197 (2008).
76. Myocardial Infarction Genetics Consortium. Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nature Genet.* **41**, 334–341 (2009).
77. Prokopenko, I. *et al.* Variants in *MTNR1B* influence fasting glucose levels. *Nature Genet.* **41**, 77–81 (2009).
78. International Schizophrenia Consortium. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* **455**, 237–241 (2008).
79. Mefford, H. C. *et al.* Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes. *N. Engl. J. Med.* **359**, 1685–1699 (2008).
80. Helbig, I. *et al.* 15q13.3 microdeletions increase risk of idiopathic generalized epilepsy. *Nature Genet.* **41**, 160–162 (2009).
81. Sharp, A. J. *et al.* A recurrent 15q13.3 microdeletion syndrome associated with mental retardation and seizures. *Nature Genet.* **40**, 322–328 (2008).
82. Bassett, A. S., Marshall, C. R., Lionel, A. C., Chow, E. W. & Scherer, S. W. Copy number variations and risk for schizophrenia in 22q11.2 deletion syndrome. *Hum. Mol. Genet.* **17**, 4045–4053 (2008).
83. McCarroll, S. A. *et al.* Deletion polymorphism upstream of *IRGM* associated with altered *IRGM* expression and Crohn's disease. *Nature Genet.* **40**, 1107–1112 (2008).
84. de Cid, R. *et al.* Deletion of the late cornified envelope *LCE3B* and *LCE3C* genes as a susceptibility factor for psoriasis. *Nature Genet.* **41**, 211–215 (2009).

Acknowledgements This paper is inspired by the deliberations of an expert working group convened by the National Human Genome Research Institute (NHGRI) on 2–3 February 2009, to address the heritability unexplained in GWAS. The authors acknowledge the participation of J. C. Cohen, M. Daly and A. P. Feinberg in the workshop.

Author Contributions T.A.M., F.S.C., N.J.C., D.B.G., L.A.H., D.J.H., M.I.M. and E.M.R. planned and participated in the workshop; L.R.C., A.C., J.H.C., A.E.G., A.K., L.K., E.M., C.N.R., M.S., D.V., A.S.W., M.B., A.G.C., E.E.E., G.G., J.L.H., T.F.C.M., S.A.M. and P.M.V. participated in the workshop; T.A.M., P.M.V., G.G., M.I.M., E.E.E., T.F.C.M. and S.A.M. drafted the manuscript; F.S.C., N.J.C., D.B.G., L.A.H., D.J.H., E.M.R., L.R.C., A.C., J.H.C., A.P.R., A.E.G., A.K., L.K., E.M., C.N.R., M.S., D.V., A.S.W., M.B., A.G.C. and J.L.H. critically reviewed and revised the manuscript for content.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/ nature. Correspondence should be addressed to T.A.M. (manoliot@mail.nih.gov).

ARTICLES

Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes

Yanli Wang¹, Stefan Juranek², Haitao Li¹, Gang Sheng¹, Greg S. Wardle², Thomas Tuschl² & Dinshaw J. Patel¹

The slicer activity of the RNA-induced silencing complex resides within its Argonaute (Ago) component, in which the PIWI domain provides the catalytic residues governing guide-strand mediated site-specific cleavage of target RNA. Here we report on structures of ternary complexes of *Thermus thermophilus* Ago catalytic mutants with 5'-phosphorylated 21-nucleotide guide DNA and complementary target RNAs of 12, 15 and 19 nucleotides in length, which define the molecular basis for Mg²⁺-facilitated site-specific cleavage of the target. We observe pivot-like domain movements within the Ago scaffold on proceeding from nucleation to propagation steps of guide–target duplex formation, with duplex zippering beyond one turn of the helix requiring the release of the 3'-end of the guide from the PAZ pocket. Cleavage assays on targets of various lengths supported this model, and sugar-phosphate-backbone-modified target strands showed the importance of structural and catalytic divalent metal ions observed in the crystal structures.

Ago is the key component of the RNA-induced silencing complex (RISC) and has an essential role in guide-strand-mediated target RNA recognition, cleavage and product release^{1–8}. Ago adopts a bilobal architecture, composed of amino-terminal PAZ-containing (N and PAZ) and carboxy-terminal PIWI-containing (Mid and PIWI) lobes. The PIWI domain adopts an RNase H fold^{9–11}, in which the catalytic Asp–Asp–Asp/His residues contribute to slicer activity^{11–13}; the Mid domain sequesters the 5'-phosphate of the guide strand^{14,15}; and the PAZ domain recognizes the 2-nucleotide overhang at the 3'-end of the guide strand^{16,17}. Ago-mediated target-RNA cleavage requires Watson–Crick pairing between guide and target, spanning both the seed segment (positions 2–8) and the cleavage site (10–11 step) as counted from the 5'-end of the guide strand^{3,4}. Endonucleolytic cleavage is mediated by Mg²⁺ cations^{18,19} and generates fragments containing a 3'-OH for the 5'-segment and a 5'-phosphate for the 3'-segment²⁰. Molecular insights into target RNA recognition and cleavage have emerged from chemical^{21,22}, biophysical²³ and structural^{5,24–27} studies, with potential application of RNA-interference-based approaches as a therapeutic modality against a range of human diseases^{28,29}.

We have previously reported on crystal structures of *T. thermophilus* Ago bound to 5'-phosphorylated 21-nucleotide guide DNA (binary complex)³⁰, and with added 20-nucleotide target RNA (ternary complex)³¹ (see Supplementary Materials for a summary of these results). A major limitation of the earlier structural study of the ternary complex³¹ was that the bases of the target RNA could not be monitored owing to disordered electron density at the 10–11 cleavage site as a result of mismatch incorporation at these steps to prevent cleavage activity. The catalytic activity of the RNase H fold of the PIWI domain of *T. thermophilus* Ago originates in Asp residues 478, 546 and 660, and hence, in the present study, single Asp to Asn, Glu or Ala mutants were incorporated at these positions to inhibit the cleavage activity. The ternary complexes of these catalytic mutants with bound guide DNA and varying target RNA lengths were then generated with 5'-phosphorylated 21-nucleotide guide DNA and fully complementary target RNAs of varying length (12, 15 and 19 nucleotides)—conditions under which both the seed segment and the cleavage site

could be potentially monitored, thereby providing insights into cleavage mechanism.

Cleavage site in Ago ternary complexes

We have solved the 2.6 Å crystal structure of the Asn 546 catalytic mutant of *T. thermophilus* Ago bound to 5'-phosphorylated 21-nucleotide guide DNA and a 12-nucleotide target RNA that is fully complementary along the length of the duplex (Fig. 1a). This is our highest resolution structure of a ternary complex to date (Fig. 1b; stereo view in a different perspective in Supplementary Fig. 1a; X-ray statistics are listed in Supplementary Table 1), and has provided detailed insights into the alignment of the guide and target strands that span both the seed segment and the cleavage site. The guide DNA strand in red can be monitored from positions 1–12 spanning the 5'-half and for positions 20–21 at the 3'-end, whereas the target RNA strand in blue can be monitored for positions 2'–12' (Fig. 1b). Both ends of the guide strand are anchored in their respective binding pockets despite formation of an 11-base-pair (bp) DNA–RNA duplex. Intermolecular contacts within the 12-nucleotide target ternary complex are highlighted in Supplementary Fig. 2. Bases 1 and 2 are splayed, with thymine at position 1 stacked over the side chain of Arg 418, and its N3 nitrogen and O4 oxygen hydrogen-bonded to the backbone (Met 413) and the side chain (Asn 436) of the Ago scaffold (Fig. 1c). Base 1 is the only residue on the guide strand that makes base-specific contacts with the Ago scaffold, and this observation is consistent with the reported sorting of small RNAs in *Arabidopsis* Ago complexes by the 5'-terminal nucleotide^{32,33}.

The guide DNA–target RNA duplex spanning positions 2 to 12 (Fig. 1d) superpositions better with an A-form helix than with its B-form counterpart (Supplementary Fig. 3a and b, respectively), with the scissile phosphate (10–11 step) on the target strand positioned opposite the catalytic residues (Asp 478, Asp 660 and Asn 546 mutant) of the RNase H fold of the PIWI domain (Fig. 1d, e). Bases 10 and 11 of the target strand stack on each other in a catalytically competent helical conformation in the ternary Ago complex (Fig. 1f), in contrast to the orthogonal arrangements of these bases owing to the insertion of Arg 548 between them in the binary Ago complex³⁰ (compare

¹Structural Biology Program, Memorial-Sloan Kettering Cancer Center, New York, New York 10065, USA. ²Howard Hughes Medical Institute, Laboratory of RNA Molecular Biology, The Rockefeller University, New York, New York 10065, USA.

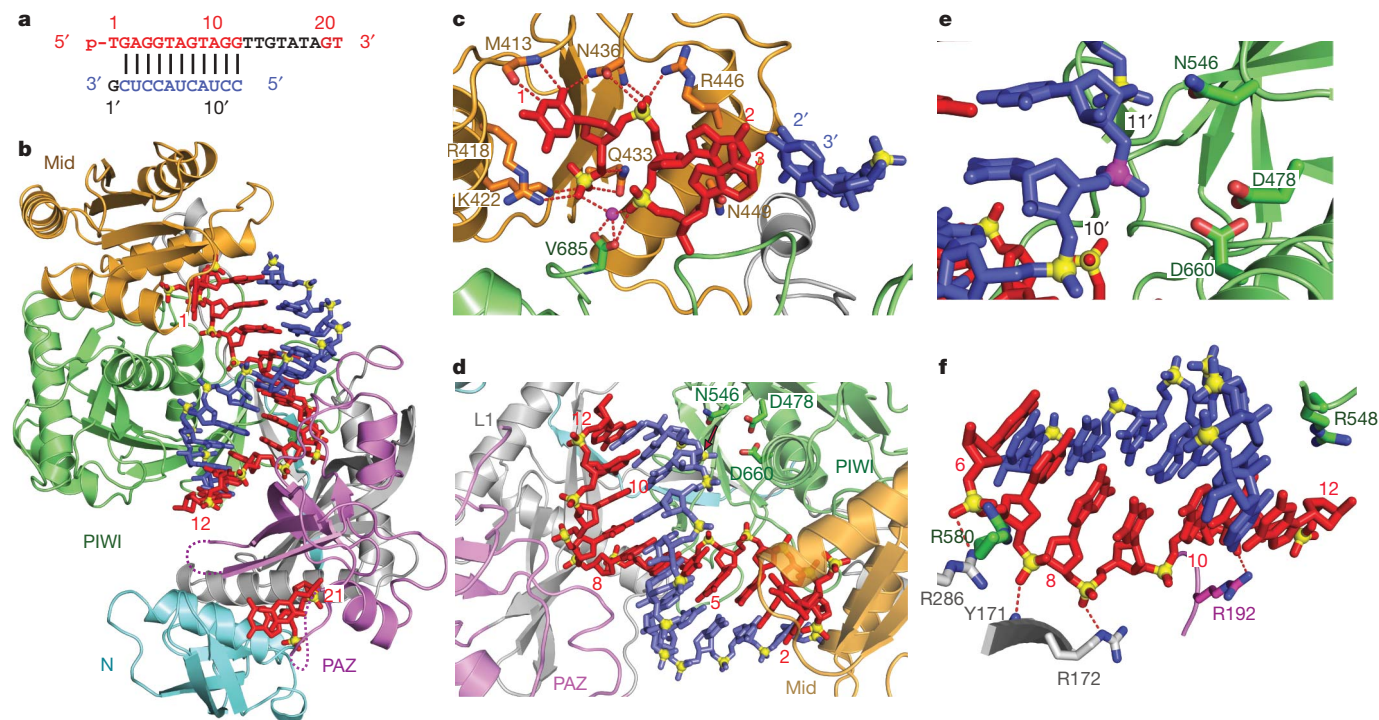


Figure 1 | Crystal structure of *T. thermophilus* Ago(Asn 546) catalytic mutant bound to 5'-phosphorylated 21-nucleotide guide DNA and 12-nucleotide target RNA. **a**, Sequence of the guide DNA–target RNA duplex. The traceable segments of the bases of the guide DNA and target RNA in the structure of the ternary complex are shown in red and blue, respectively. Disordered segments of the bases on both strands that cannot be traced are shown in grey. **b**, View of the 2.6 Å crystal structure of the Ago ternary complex. The Ago protein domains (N in cyan, PAZ in magenta, Mid in orange, PIWI in green) and linkers (L1 and L2 in grey) are colour-coded. The bound 21-nucleotide guide DNA (red) is traced for bases 1–12 and 20–21, whereas the bound 12-nucleotide target RNA (blue) is traced for bases 2'–12'. Backbone phosphorus atoms are yellow. Both ends of the bound guide DNA are anchored. **c**, Expanded view of the ternary complex highlighting the alignment of guide DNA (1–3) and target RNA (2'–3'), where the bases of the 1–2 step of the guide strand are splayed. Note the intermolecular hydrogen-bonding of the Watson–Crick edge of T1 with the

backbone amide carbonyl of Met 413 and side chain of Asn 436, as well as the positioning of phosphate 1 of the guide strand in the Mid binding pocket. A Mg^{2+} cation (purple) coordinates to phosphates 1 and 3 of the guide strand, as well as to an inserted carboxylate of Val 685 from the C terminus.

d, Expanded view of the ternary complex highlighting the guide DNA (1–12)–target RNA (2'–12') duplex, together with the catalytic residues (Asp 478, Asp 660 and Asn 546 mutant) of the RNase H fold of the PIWI domain. The scissile phosphate group at the 10'–11' step of the target RNA is indicated by a red arrow. **e**, Expanded view highlighting the positioning of the backbone phosphate linking the 10'–11' step (phosphorus coloured in magenta) of the target RNA relative to the catalytic residues (Asp 478, Asp 660 and Asn 546 mutant) in the ternary complex. **f**, Positioning of the side chain of Arg 548 relative to the guide DNA (6–12)–target RNA (6'–12') duplex. Note the intermolecular contacts between the sugar–phosphate backbone of the guide strand and side chains of the protein in the ternary complex.

Supplementary Fig. 4a (binary) with 4b (ternary)). Conformational changes in both the guide strand (Supplementary Fig. 5a) and Ago (Supplementary Fig. 5b) accompany the transition from binary to ternary complex formation (Supplementary Fig. 6 and Supplementary Movie 1).

Release of guide 3'-end from PAZ pocket

Next we solved the 3.05 Å crystal structure of the Glu 546 catalytic mutant of *T. thermophilus* Ago bound to 5'-phosphorylated 21-nucleotide guide DNA and a 15-nucleotide target RNA that is fully complementary along the length of the duplex (Fig. 2a; stereo view in Supplementary Fig. 1b; X-ray statistics are listed in Supplementary Table 1). The guide DNA strand can be monitored from positions 1–16, whereas the target RNA strand can be monitored from positions 2'–15' (Fig. 2b). The 5'-phosphate of the guide strand is still anchored in the Mid pocket, but the 3'-end (positions 17–21) is disordered and cannot be traced) is released from the PAZ pocket on formation of the 14-bp duplex spanning positions 2–15 of the guide strand. The molecular basis for the release of the 3'-end of the guide strand is that the helical conformation for nucleotides 12–15 disallows the 3'-end from reaching the binding pocket in the PAZ domain.

We observe conformational changes on proceeding from the ternary Ago complex with bound 12-nucleotide target (Fig. 2c) to its counterpart with bound 15-nucleotide target (Fig. 2d), and these changes can

be visualized after superpositioning of the PIWI-containing (Mid and PIWI) lobe as shown by the yellow arrow in Fig. 2e (also see Supplementary Movie 2). These changes involve a pivotal rotation of the PAZ domain (compare PAZ domain alignments in Supplementary Fig. 7a and b), as well as movement of loops L1 and L2 located on the nucleic-acid-interfacing surface of the PIWI domain (Fig. 2f).

Details of intermolecular contacts between loop L1 and the guide DNA 11–12 segment in the 12-nucleotide target RNA ternary complex are shown in Fig. 2g, whereas intermolecular contacts between loops L1 and L2 and the guide DNA 11–15 segment in the 15-nucleotide target RNA ternary complex are shown in Fig. 2h. Notably, L1 changes from a loop (Fig. 2g) to a β -turn (Fig. 2h) on proceeding from the 12- to the 15-nucleotide target RNA ternary complexes, resulting in several extra hydrogen bonds within this β -turn and with loop 2, thereby stabilizing this new conformation. The conformational transitions in loops L1 and L2 are required to avoid steric clashes with the DNA guide strand (Supplementary Fig. 8) on addition of three more base pairs on proceeding from the 12- to the 15-nucleotide target RNA ternary complexes. Unexpectedly, changes in the conformation of loop L1 force the attached β -strand encompassing residues 489–493, as part of a multi-stranded β -sheet, to slide by a single residue with the accompanying flip of the entire β -strand and its side chains, on proceeding from the 12- to the 15-nucleotide target ternary complex (Fig. 2i, identified by a black double-edged arrow in Fig. 2f and Supplementary Fig. 9).

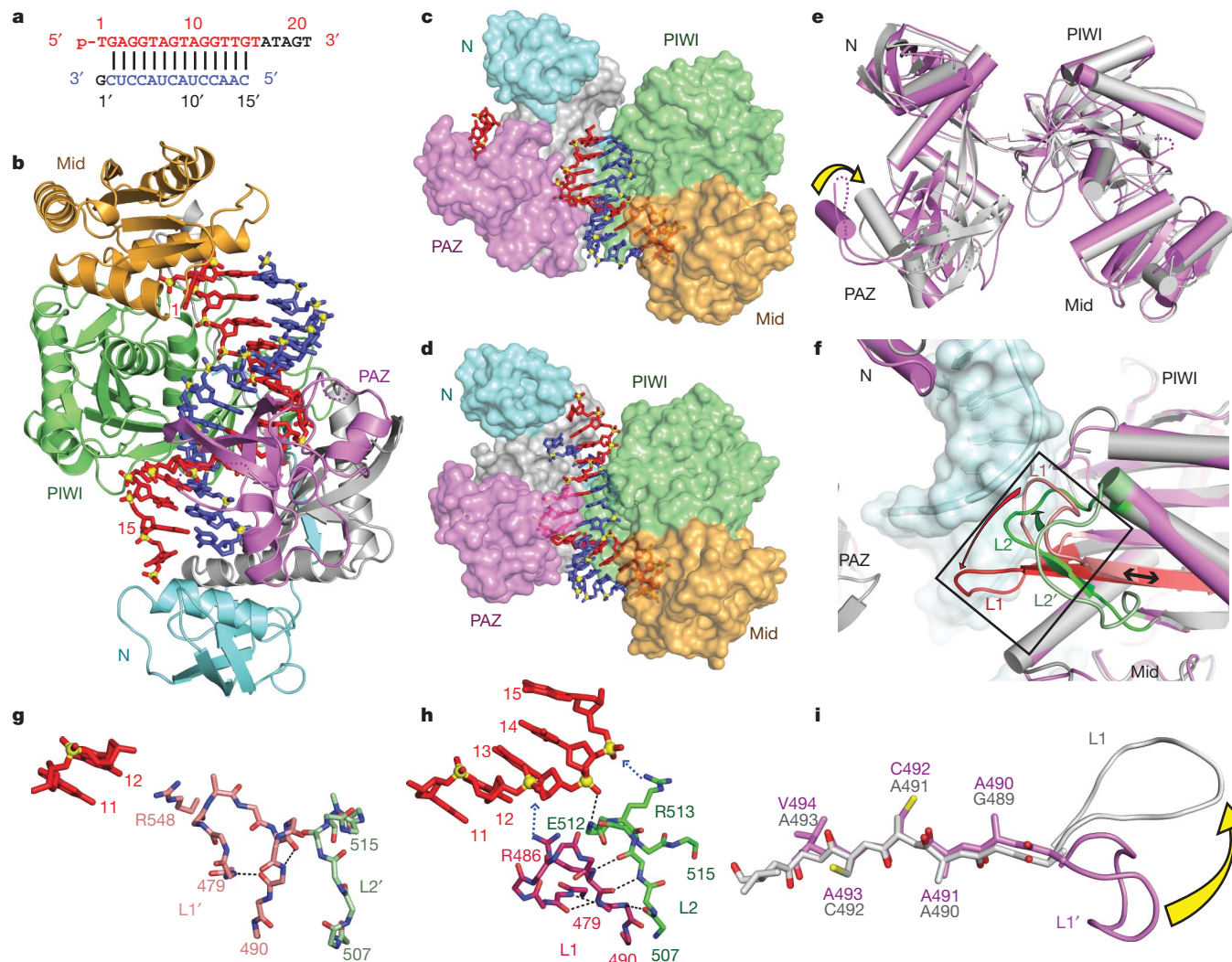


Figure 2 | Crystal structure of *T. thermophilus* Ago(Glu 546) catalytic mutant bound to 5'-phosphorylated 21-nucleotide guide DNA and 15-nucleotide target RNA. **a**, Sequence of the guide DNA–target RNA duplex, with traceable segments colour-coded as in Fig. 1a. **b**, View of the 3.05 Å crystal structure of the Ago ternary complex, colour-coded as outlined in Fig. 1b. The bound 21-nucleotide guide DNA (red) is traced for bases 1–16, whereas the bound 15-nucleotide target RNA (blue) is traced for bases 2'–15'. Only the 5'-end of the guide DNA is anchored in this ternary complex. **c**, **d**, Comparison of the crystal structures of mutant Ago(Asn 546)–12-nucleotide target (**c**) and of mutant Ago(Glu 546)–15-nucleotide target (**d**) ternary complexes. The Ago protein is shown in a surface representation with domains and linkers colour-coded as in Fig. 1b. The guide DNA (red) and target RNA (blue) are shown in stick representation with backbone phosphorus atoms in yellow. **e**, View of the alignment of mutant Ago(Asn 546)–12-nucleotide target complex (magenta) and mutant Ago(Glu 546)–15-nucleotide target complex (silver), after superpositioning of their PIWI-containing (Mid and PIWI) modules. The yellow arrow indicates the magnitude of the conformational change on proceeding from the 12-nucleotide target to 15-nucleotide target ternary complexes. **f**, Conformational changes in loop 1 (residues 479–488, red arrow) and loop 2 (residues 505–516, green arrow) of the PIWI domain on

In mechanistic terms, we favour the view that the conformational transitions in loops L1 and L2 and associated sliding and flipping of the β -strand are triggered by widening of the substrate-binding channel between the PIWI and N domains to accommodate a lengthening of the A-form duplex from 11-bp in the 12-nucleotide target RNA complex to 14-bp in the 15-nucleotide target RNA complex. Such changes not only push the PAZ domain away but also release the 3' end of guide strand from the PAZ-binding pocket (Figs 1b, 2b and Supplementary Fig. 7). Moreover, we note that sliding and flipping of the β -strand

proceeding from the 12-nucleotide target ternary complex (magenta) to the 15-nucleotide target ternary complex (silver). Only the DNA–RNA duplex for the 15-nucleotide target ternary complex is shown in cyan in a surface representation. Loops 1 and 2 are coloured light red (labelled L1') and light green (labelled L2') in the 12-nucleotide target ternary complex, and dark red (labelled L1) and dark green (labelled L2) in the 15-nucleotide target ternary complex. The β -strand involved in sliding is highlighted by a black double-edged arrow. **g**, Ternary complex containing 12-nucleotide target RNA. Residues 11 and 12 of the guide strand are in red, and loops L1' and L2' are in light red and light green, respectively. **h**, Ternary complex containing 15-nucleotide target RNA. Residues 11 to 15 of the guide strand are in red, and loops L1 and L2 are in dark red and dark green, respectively. Loop L1 switches to a β -turn aligned by hydrogen bonding within the turn and also with loop L2, thereby stabilizing this turn conformation. The main-chain of Glu 512 forms a hydrogen bond with the phosphate group of residue 14 of the guide DNA. The positively charged side chains of Arg 513 and Arg 486 interact with the backbone of the DNA guide strand, as indicated by blue arrows. **i**, Ribbon representation of the sliding of the β -strand (Gly 489 to Val 494) by one residue, and conformational transition in adjacent L1 loop on proceeding from the 12-nucleotide target RNA ternary complex (magenta) to 15-nucleotide target RNA ternary complex (silver).

occurs with minimal perturbation of β -sheet formation (schematic in Supplementary Fig. 9), and flipping of the entire β -strand does not disrupt specific side-chain interactions.

We have compared the structures of Ago mutant ternary complexes with 12-nucleotide (Fig. 1b) and 15-nucleotide (Fig. 2b) target RNAs reported in this study with the previously reported structure of the ternary complex of wild-type Ago with 20-nucleotide target RNA containing a pair of mismatches at the cleavage site³¹. The previous structure of the ternary complex (two molecules in the asymmetric

unit)³¹ and one solved recently in a different crystal form (one molecule in the asymmetric unit; X-ray crystallographic statistics in Supplementary Table 2) in which segment 2–9 is fully paired and both ends of the guide strand are anchored, are most similar to the ternary complex with 12-nucleotide target RNA in the present study, in which segment 2–12 is fully paired and both ends of the guide strand are also anchored (comparison outlined in Supplementary Fig. 10a, b).

Our studies resolve a mechanistic issue related to guide-strand-mediated recognition and cleavage of target RNA within Ago complexes. Several groups have proposed a ‘two-state’ model in which the guide strand is anchored at both of its ends during the nucleation step of target recognition, but its 3′-end is released from the PAZ pocket owing to topological constraints, after propagation of the duplex towards the 3′-end of the guide strand^{4,11,34}. An alternative ‘fixed-end’ model proposed that both ends of the guide strand remain anchored during the nucleation and the propagation steps of RNA recognition³⁴. Our results support a two-state mechanism for the system under study, given that our structures demonstrate that both ends of the guide strand are anchored in a ternary complex containing one turn of the A-form helix (12-nucleotide target RNA) spanning the seed segment and cleavage site (Fig. 1b), but the 3′-end is released from the PAZ pocket on extending this duplex by three more base pairs (15-nucleotide target RNA) towards the 3′-end of the guide strand (Fig. 2b).

N domain blocks guide–target pairing beyond position 16

The 2.8 Å crystal structure of the Asn478 catalytic mutant of *T. thermophilus* Ago bound to 5′-phosphorylated 21-nucleotide guide DNA and a 19-nucleotide target RNA (sequence in Fig. 3a, structure in Fig. 3b, stereo view in Supplementary Fig. 1c; X-ray statistics are listed in Supplementary Table 1) is similar to the Ago(Glu 546) catalytic mutant ternary complex with 15-nucleotide target RNA (Fig. 2b), except that one extra base pair can be traced, allowing monitoring of 15-bp of guide–target duplex spanning positions 2–16 of the guide strand (stereo electron density maps of the guide and target strands are shown in Supplementary Fig. 11). Intermolecular contacts within the 19-nucleotide target ternary complex are highlighted in Supplementary Fig. 12). Furthermore, the sugar-phosphate backbone of the target strand is intact at the 10–11 step, and on either side of it, for both 15- and 19-nucleotide target ternary complexes (see $F_o - F_c$ omit maps contoured at 3.7σ in Supplementary Fig. 13a and b, respectively).

An unexpected mechanistic insight to emerge from our structural studies of the three ternary Ago complexes outlined earlier is that the guide DNA–target RNA duplex retains the A-form duplex architecture spanning the seed segment, the cleavage site and observable elements towards the 3′-end of the guide strand (up to position 16), and it is solely the Ago scaffold that adjusts by pivot-like domain movements, to relieve the topological stress associated with zippering up the RNA target through pairing with its guide-strand template. A

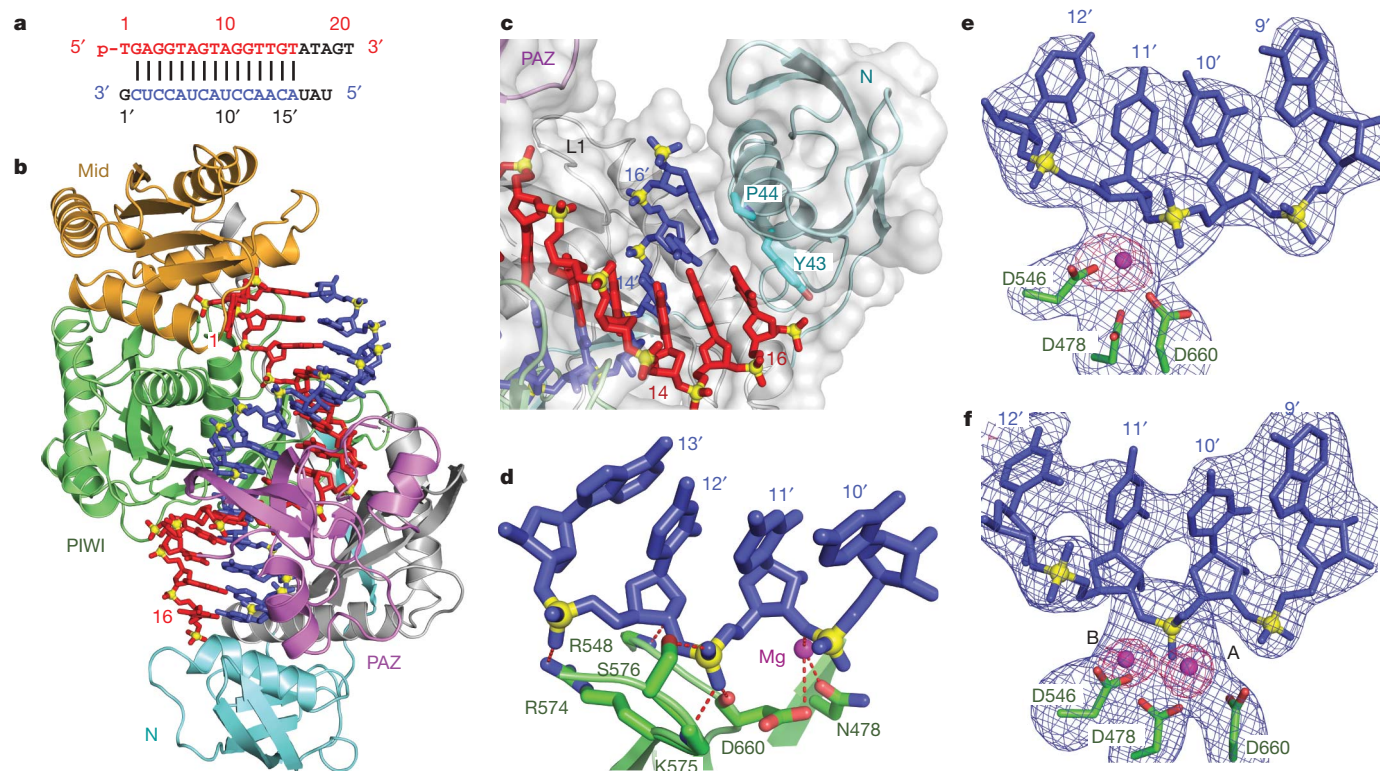


Figure 3 | Crystal structure of *T. thermophilus* Ago(Asn 478) catalytic mutant bound to 5′-phosphorylated 21-nucleotide guide DNA and 19-nucleotide target RNA and identification of Mg^{2+} binding sites within the catalytic pocket of the wild-type Ago complex. **a**, Sequence of the guide DNA–target RNA duplex, with traceable segments colour-coded as in Fig. 1a. **b**, View of the 2.8 Å crystal structure of the ternary complex, colour-coded as outlined in Fig. 1b. The bound 21-nucleotide guide DNA (red) is traced for bases 1–16, whereas the bound 19-nucleotide target RNA (blue) is traced for bases 2′–16′. Only the 5′-end of the guide strand is anchored in this ternary complex. **c**, Expanded view of the 19-nucleotide target ternary complex highlighting blocking of propagation of the guide DNA–target RNA duplex beyond pair 16 by the N domain. Base 16 of the guide strand stacks over the aromatic ring of Tyr 43, whereas base 16′ of the target strand stacks over

Pro 44. **d**, Intermolecular hydrogen-bonding contacts between the sugar-phosphate backbone of the 10′–13′ target RNA segment and backbone and side chains of the PIWI domain in the 19-nucleotide target ternary complex. **e**, **f**, $F_o - F_c$ omit maps (blue colour, contoured at 3.5σ) of the 9′–12′ segment of bound RNA and catalytic Asp 478, Asp 546 and Asp 660 residues in the 3.3 Å structures of the ternary complexes in 50 mM Mg^{2+} (**e**, space group $P4_32_12_1$, one molecule in the asymmetric unit) and in 80 mM Mg^{2+} (**f**, space group $P2_12_12_1$, two molecules in asymmetric unit). Bound Mg^{2+} cation(s) were identified in omit maps contoured in purple at 6.0σ as outlined in **e** and **f**, based on coordination to several oxygen atoms in an approximate octahedral geometry. One bound Mg^{2+} cation can be assigned in the ternary complex in 50 mM Mg^{2+} in **e**, and two bound Mg^{2+} cations can be assigned in the ternary complex in 80 mM Mg^{2+} in **f**.

second unanticipated observation is that the N domain blocks propagation of the guide DNA–target RNA duplex beyond position 16 in the 19-nucleotide target ternary complex (Fig. 3c), with the base at position 16 of the guide strand stacking on the aromatic ring of Tyr 43, and the base at 16' of the target strand stacked over the Pro 44 ring. Thus, base pairing is disrupted for steps 17, 18 and 19, with anticipated trajectories for the separated guide and target strands schematized in Supplementary Fig. 14.

The sugar-phosphate backbone spanning the seed segment of the guide but not the target strand is hydrogen-bonded to the protein (see Supplementary Movie 3). We also note that the sugar-phosphate backbone of the target RNA spanning the 10'–13' segment forms intermolecular hydrogen bonds with the Ago scaffold in the 12-nucleotide (Supplementary Fig. 15a), 15-nucleotide (Supplementary Fig. 15b) and 19-nucleotide (Fig. 3d) target ternary complexes, establishing the potential for photochemically facilitated cross links between this segment of the target RNA and its spatially identified proximal sites on the protein³⁵.

A pair of Mg^{2+} cations mediates cleavage chemistry

The PIWI domain of Ago adopts an RNase H fold^{9–11,30,31}, with catalytic Asp 478, Asp 546 and Asp 660 residues lining the active site of the *T. thermophilus* enzyme. Two Mg^{2+} cations have been shown to facilitate RNA hydrolysis during catalytic cleavage by RNase-H-containing nucleases, with cation A assisting nucleophilic attack by positioning and activating a water molecule, and cation B stabilizing the transition state and leaving group^{36,37}. Because catalytic mutations could induce distortions of the optimal geometry for coordination to divalent cations, we attempted to identify bound Mg^{2+} cation(s) in the catalytic pocket of the ternary complex of wild-type *T. thermophilus* Ago with 19-nucleotide target RNA, that is fully complementary to positions 2–19 of the guide strand (Fig. 3a).

Crystals of the Ago ternary complex were grown as a function of Mg^{2+} concentration, with 3.3 Å data sets collected for crystals in 50 mM Mg^{2+} (space group $P4_32_12$, one molecule in the asymmetric unit) and 80 mM Mg^{2+} (space group $P2_12_12_1$, two molecules in asymmetric unit) solution (X-ray statistics listed in Supplementary Table 3). Gel electrophoresis of the crystals established that the target RNA was not cleaved in either complex, presumably because *T. thermophilus* Ago-mediated cleavage is optimal at higher temperatures and has a marked preference for Mn^{2+} over Mg^{2+} (ref. 11). The $F_o - F_c$ omit maps (blue colour, contoured at 3.5σ) of the target strand residues 9'–12' and catalytic Asp residues for the Ago ternary structures in 50 mM Mg^{2+} and 80 mM Mg^{2+} are shown in Fig. 3e and f, respectively. A single bound Mg^{2+} , positioned towards the leaving group side of the scissile phosphate (cation B) can be identified in the structure in 50 mM Mg^{2+} (Fig. 3e, omit map contoured in purple at 6.0σ), with an intact target RNA readily traceable for the 9'–12' segment. A pair of Mg^{2+} cations separated by 3.9 Å, which coordinate the hydrolysis of the scissile phosphate, were identified in the structure in 80 mM Mg^{2+} (Fig. 3f). The assignment of the extra density to Mg^{2+} site(s) at 3.3 Å resolution is based on coordination of the divalent cation(s) to several oxygen atoms in an approximate octahedral geometry (stereo views in Supplementary Fig. 16a, b). Of the three catalytic Asp residues lining the catalytic pocket, only Asp 478 coordinates to both Mg^{2+} cations (Fig. 3f and Supplementary Fig. 16b). The structures of the catalytic residues, Mg^{2+} sites and RNA backbone for *B. halodurans* RNase H (1.85 Å) and *T. thermophilus* Ago (3.3 Å) complexes are superpositioned in stereo for comparative purposes in Supplementary Fig. 17. Given that the crystals of the ternary complexes grown from both 50 and 80 mM Mg^{2+} diffract to 3.3 Å resolution, it is at present not possible to identify the position of the water molecule that would participate and be positioned for in-line attack on the scissile phosphate.

We observe detectable conformational changes after superpositioning of the single and the pair of Mg^{2+} -bound ternary complex structures through their PIWI-containing lobes. These changes

are restricted to the PAZ domain (Supplementary Fig. 18a) and the target RNA strand (Supplementary Fig. 18b). The catalytic residues are optimally positioned for cleavage in the structure of the ternary complex with a pair of Mg^{2+} cations.

Thus, the Ago protein, capitalizing on the RNase H fold of its PIWI domain^{9–11}, uses three catalytic Asp residues and two Mg^{2+} cations to facilitate site-specific cleavage of RNA targets, yielding products containing 5'-phosphate and 3'-OH ends²⁰, a feature in common with members of the retroviral integrase superfamily³⁷.

Analysis of the catalytic activity of *T. thermophilus* Ago

Target RNA cleaving bacterial complexes are most effectively reconstituted using single-stranded guide DNA rather than RNA^{11,15,30,31}. To explore whether DNA might also function as a target, we subjected chemically synthesized DNA and RNA targets (Supplementary Table 4) to DNA-guided Ago cleavage reactions. DNA is resistant to hydrolysis by divalent metal ions and high temperature incubation, thereby yielding a clearer picture of target cleavage. *T. thermophilus* Ago loaded with guide DNA derived from luciferase sequence studied previously^{30,31} cleaved DNA as well as RNA targets; however, several unexpected minor cleavage products were also observed (Supplementary Fig. 19). These side products resulted from partial self-complementarity of the guide DNA, leading to cleavage of guide DNA during the Ago loading process and acceptance of the shorter cleavage products as guide DNAs. We therefore tested new guide and target sequence pairs, identical to the microRNA let-7 sequence selected for crystallography. The let-7 guide and target molecules yielded a single cleavage band, with DNA being a better substrate than RNA (Supplementary Fig. 20a). Target DNA cleavage occurred in the presence of Mg^{2+} or Mn^{2+} , but not Ca^{2+} (Supplementary Fig. 20b), supporting single and multiple turnover (Supplementary Fig. 20c). Cleavage products started to accumulate after a short (about 2 min) lag phase, at an approximate rate constant of 0.1 min^{-1} under single turnover (0.5 μM target) and $0.2\text{--}0.4 \text{ min}^{-1}$ under multiple turnover (5 μM target) conditions (Supplementary Fig. 20c). These rate constants indicate that our cleavage conditions are approaching substrate saturation and that product release is not rate limiting. We also included cleavage experiments using mutant Ago proteins that were used for the crystal structures (Figs 1–3) and tested for DNA-guided RNA (Supplementary Fig. 21a) or DNA (Supplementary Fig. 21b) target cleavage. Of the mutant Agos, only the Asn 546 mutant showed some residual activity, and product formation was reduced >500-fold.

Minimal target DNA requirements

Previously, we showed that luciferase guide DNA strands as short as 9 nucleotides promoted target RNA cleavage; the minimal target length was not addressed³¹. We first shortened the let-7 DNA target (Fig. 4a) from its 5' end (Fig. 4b). Truncation of the target to 16 nucleotides did not alter cleavage activity, but 15- and 14-nucleotide targets showed 120- and 400-fold reduced cleavage rates, respectively, and a 12-nucleotide target was not cleaved. This indicates that residues 17' and higher do not contribute to cleavage, and was further supported by our finding that 21- or 24-nucleotide DNA targets, in which regions 17'–21' or 17'–24' were unpaired with same size guides, showed similar activity compared to their fully paired versions (Supplementary Fig. 22).

To examine the importance of the 3' end of the target, we tested 15-nucleotide DNA target strands displaced in 1-nucleotide steps relative to the let-7 target (Fig. 4c). DNA targets covering 2'–16', 3'–17' and 4'–18' showed cleavage activity similar or better than 21-nucleotide-long targets, but 100- and 500-fold reduced rates were obtained for targets covering 5'–19' and 6'–20'. These experiments indicate that positions 1' to 3' were dispensable for target cleavage.

In summary, positions 4' to 16' need to be paired to facilitate efficient target DNA cleavage when presented to *T. thermophilus* Ago loaded with 21-nucleotide guide DNA. On the other hand, guide

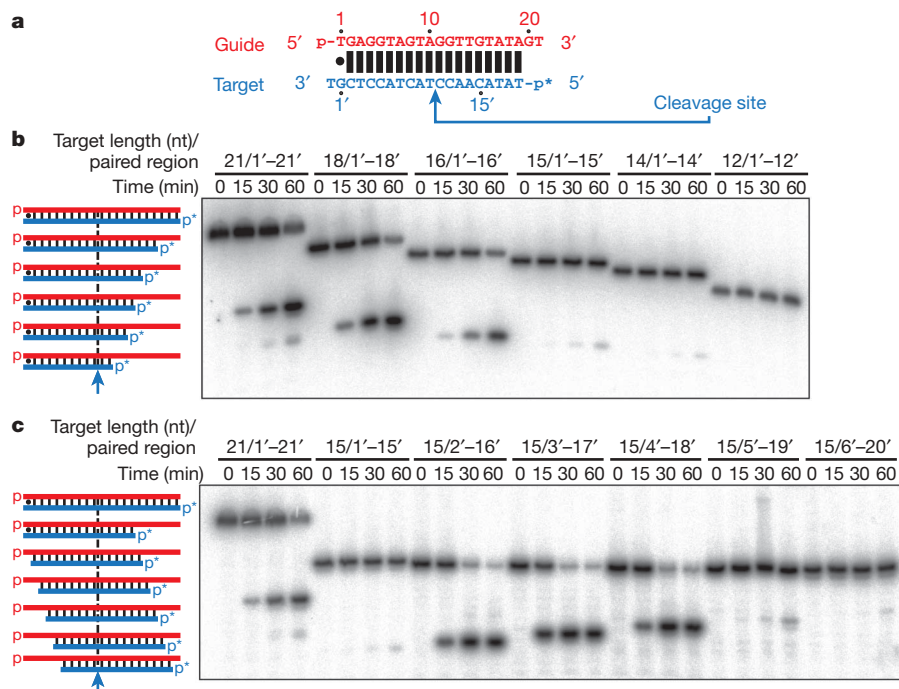


Figure 4 | Effect of complementarity and length on target DNA cleavage by *T. thermophilus* Ago. Cleavage reactions were performed as described in the Methods, and products were resolved on denaturing polyacrylamide gels; for DNA sequences, see Supplementary Table 4. **a**, Schematic of the reference DNA duplex utilized for length variation experiments; the cleavage site is indicated by an arrow, the position of the ^{32}P label by an asterisk. **b**, Shortening of the target DNA from its 5' end. Alterations of the target DNA and corresponding paired structure are illustrated to the left. Target DNA cleavage was performed at 65 °C rather than 75 °C to facilitate hybridization of shortened targets. nt, nucleotides. **c**, Positional variation of 15-nucleotide target DNAs. For labelling and reaction conditions, see **b**.

DNA as short as 9 nucleotides promoted *T. thermophilus* Ago cleavage of target RNA, indicating that base-pairing involving residues 10' to 16' *per se* is not essential. Short guides, in contrast to 21-nucleotide guides, are unable to occupy the PAZ domain with their 3' ends. Therefore, we speculate that transitioning of the Ago ternary complex into a cleavage-active conformation requires either the release of the guide 3'-end PAZ interaction or its initial absence as seen for short guide strands. Release of the PAZ guide 3'-end interactions is driven by base-pairing including position 16' of a target.

It may seem surprising that the Ago conformation of the 15- and 19-nucleotide target-RNA-containing structures were similar. However, the thermodynamic stability of DNA–RNA duplexes is different from DNA–DNA duplexes³⁸, and fewer but more stable base pairs may facilitate the switch to the active conformation. In support of this view, we observed that the cleavage activity for the 15-nucleotide (positions 1'–15') and a 16-nucleotide (positions 1'–16') target RNAs (Supplementary Fig. 23) only differed by 1.4-fold and was comparable to that of the longer target RNA (Supplementary Fig. 21).

Our crystal structures also indicated that base pairs involving positions 17' or higher could not form owing to steric clashes with the N-terminal domain. To test whether propagation of the duplex beyond position 16' could contribute to catalysis, we tested Ago deletion mutants del(1–106) and del(1–177) but found that they lost all activity (Supplementary Fig. 21a, b). This suggests that the N domain also has a crucial involvement in transitioning or stabilizing the active conformation of the ternary complex, and could possibly even affect other steps including loading of the guide DNA, which were not tested.

Target DNA sugar-phosphate backbone role during cleavage

To assess the contribution of sugar and phosphate residues during target DNA recognition and cleavage, we introduced 2'-hydroxyl (OH) and 2'-methoxy (Ome) modifications at positions 9', 10' or 11', as well as 2'-fluoro (F) at positions 10' or 11' (Fig. 5a, Supplementary Fig. 24 and Supplementary Table 4). OH, Ome and F

2'-modified ribonucleosides favour the A-helical C3'-endo ribose conformation, whereas deoxynucleotides are preferable in the B-helical C2'-endo conformation³⁹, and therefore stabilize double-helical structures. The most profound effects on cleavage were shown by 2'-substitutions at residue 11', which are immediately adjacent to the cleaved phosphodiester bond. The 2'-F substitution enhanced the single (Fig. 5a) and multiple (Supplementary Fig. 24) turnover cleavage rate by approximately 4- and 6-fold, respectively, compared to 2'-H, presumably because the electronegative 2'-F group is able to stabilize the developing negative charge of the 3' oxygen leaving group during the transition state. The cleavage rate was reduced twofold by 2'-OH at residue 11', and 2'-Ome completely abrogated cleavage, presumably by affecting the hydration pattern optimal for stabilization of the transition state. Also, there is no evidence for hydrogen bonding of the 2' residue to neighbouring nucleotides or amino-acid side chains. Taken together, the drastic effects on reaction rates by 2' modifications at the 11' position cannot be rationalized by simple differences in sugar conformation, but by a combination of electronic and steric effects differentially affecting the transition state. Modifications of the 2' position one nucleotide removed from the cleavage site showed less or no effect; in contrast, position 9' showed an unanticipated threefold reduction in rate for 2'-OH and 2'-Ome (Fig. 5a).

To probe the role of phosphate oxygens, which can coordinate structurally or catalytically important divalent metal ions⁴⁰, we synthesized the mixed phosphorothioate diastereomers located between residues 8' and 9', 9' and 10', 10' and 11', or 11' and 12', and purified by reverse-phase high-performance liquid chromatography (HPLC) the *S_p* form to >85%, and the *R_p* form to >97% purity. Cleavage reactions were performed in the presence of either 5 mM Mg^{2+} , which preferably coordinates to oxygen, or 5 mM Mn^{2+} , which preferably coordinates to sulphur. Phosphorothioate substitution at the cleavage site, positions 10'–11', showed the most profound effects (Fig. 5b). In Mg^{2+} -containing buffer, the *S_p* form was inactive and the *R_p* form was reduced 200-fold in single-turnover cleavage rates. The loss of activity

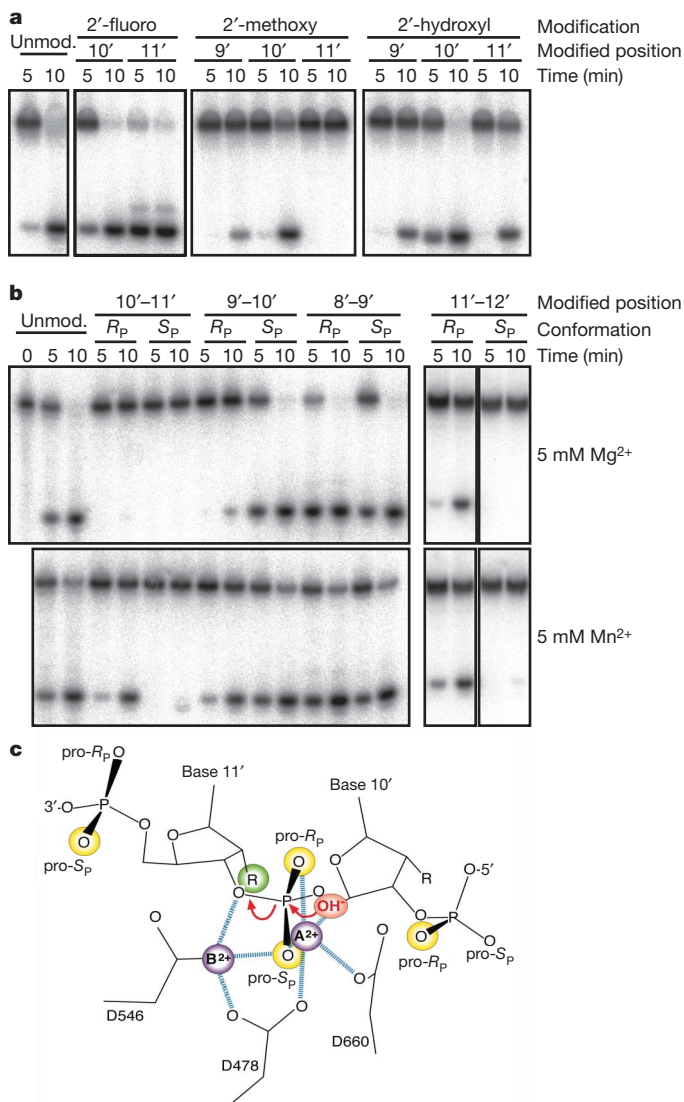


Figure 5 | Effect of sugar-phosphate backbone modifications on target DNA cleavage by *T. thermophilus* Ago. Cleavage experiments were performed as described in Methods. **a**, 2'-fluoro-, 2'-methoxy- and 2'-hydroxyl-substitutions of single 2'-deoxyribose residues of the target DNA strand at and near the cleavage site. The control target (unmod.) was the unmodified oligodeoxynucleotide. **b**, Phosphorothioate modification of the target DNA. The phosphate configuration (R_p or S_p) of the phosphorothioate diastereomers is indicated. Cleavage assays were performed in the presence of either Mg²⁺ or Mn²⁺ cations. Note that the experiment for the 11'-12' isomers was a different experiment, in which overall reaction rates were slower. For the complete experiment see Supplementary Fig. 25. Sequences of oligonucleotides are in Supplementary Table 4. **c**, Structure of the cleavage site modelling the attack of the hydroxyl nucleophile. Phosphate oxygen and active site carboxylate oxygens coordinated to metal ions A and B (purple spheres), with distances less than 2.5 Å shown as blue dashed lines. The coordination of the carboxylate oxygen from Asp 546 to metal ion B is hidden in the projection. The phosphate oxygens and 2' residues sensitive to modification are shown as yellow and green spheres, respectively; R denotes 2'-H, -OH, -F or -Ome. Red arrows indicate the attack of the hydroxyl nucleophile modelled to be directly coordinated by metal ion A, and the stabilization of the developing negative charge of the 3' oxyanion leaving group by metal ion B.

of the R_p form was rescued by Mn²⁺, yielding a less than twofold reduction compared to 2'-H; however, the S_p form remained inactive. Phosphorothioate substitutions more distant to the cleavage site either had no effect (R_p and S_p at positions 8'-9', S_p at positions 9'-10' and R_p at positions 11'-12'), or were reduced by 15-fold and by more than 80-fold for R_p at positions 9'-10' and S_p at positions

11'-12', respectively, and rescued by Mn²⁺ to less than twofold and more than 20-fold, respectively. Non-bridging phosphate oxygens that are sensitive to sulphur substitution and responsive to Mn²⁺ rescue are believed to directly coordinate to Mg²⁺, and the interaction stabilizes ground and transition states of the cleavage reaction to a similar degree. A phosphate oxygen sensitive to phosphorothioate substitution, but without metal ion rescue feature, such as the pro-S_p oxygen at the cleavage site, probably differentially stabilizes the transition state versus the ground state. Substituting the 10'-11' pro-S_p oxygen by sulphur increases the bond length by about 0.6 Å—distance sufficient to perturb the complex network of interactions coordinated at this phosphate oxygen (Fig. 5c). The pro-S_p oxygen is coordinated to metal ions A and B, with A positioning the attacking hydroxyl ion nucleophile and B stabilizing the leaving 3' oxyanion. The importance of stabilizing the leaving group was also documented earlier by the effects of modifications at the adjacent 2' position. In contrast, the pro-R_p oxygen at the cleavage site is only coordinated to metal ion A, and the sulphur substitution was rescued with Mn²⁺, indicating more flexibility for positioning the nucleophile by metal ion A.

Structural overview and functional implications

Our current structures of ternary complexes with catalytic mutants of *T. thermophilus* Ago have defined the positioning of the guide DNA-target RNA A-form duplex relative to the catalytic Asp residues of the RNase H fold of the PIWI domain, thereby establishing the molecular basis for site-specific cleavage at the phosphate bridging the 10'-11' step of the target strand. Further structural studies of ternary complexes with wild-type Ago have identified two Mg²⁺ cations within the catalytic pocket, located on either side of the cleavable phosphate, thereby positioned to mediate the cleavage chemistry. Both ends of the guide strand are anchored in the ternary complex composed of one turn of the DNA-RNA duplex spanning the seed segment and cleavage site, but consistent with a two-state model, the 3'-end is released from the PAZ pocket after propagation of the guide-target duplex by three additional base pairs. Notably, the guide DNA and target RNA form a regular A-form helix spanning a maximum of 15 base pairs (positions 2-16), with the Ago scaffold undergoing pivot-like domain movements as the target RNA zippers up by pairing with its guide template.

The kinetic effects of target site phosphorothioate substitution and 2' modification during Ago-mediated DNA cleavage are rationalized by the crystal structure, and consistent with the mechanism of RNase H cleavage studied in other systems³⁷. The absence of amino-acid side chains able to interrogate whether the target presented at the active site is RNA or DNA might suggest that DNA could be a more probable target of this bacterial Ago protein, as seen for other members of the retroviral integrase superfamily in which Ago proteins belong³⁷.

METHODS SUMMARY

Wild-type and mutant *T. thermophilus* Ago proteins were overexpressed from *Escherichia coli* and purified by chromatography as described previously³⁰. Crystals were obtained by the hanging-drop or sitting-drop vapour diffusion. The ternary Ago complex was generated in a stepwise manner by initially mixing the protein with 5'-phosphorylated 21-nucleotide guide DNA, followed by addition of different length target RNAs. All wild-type and mutant Ago complex structures were determined by molecular replacement using the domains of the binary Ago complex structure (Protein Data Bank accession code 3DLH)³⁰ as search models. Cleavage assays were undertaken with let-7 guide and target oligonucleotides. Details of all crystallographic and biochemical procedures are listed in Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 7 May; accepted 18 August 2009.

1. Dykxhoorn, D. M., Novina, C. D. & Sharp, P. A. Killing the messenger: short RNAs that silence gene expression. *Nature Rev. Mol. Cell Biol.* **4**, 457-467 (2003).
2. Baulcombe, D. RNA silencing in plants. *Nature* **431**, 356-363 (2004).

3. Meister, G. & Tuschl, T. Mechanisms of gene silencing by double-stranded RNA. *Nature* **431**, 343–349 (2004).
4. Tomari, Y. & Zamore, P. D. Perspective: machines for RNAi. *Genes Dev.* **19**, 517–529 (2005).
5. Filipowicz, W., Jaskiewicz, L., Kolb, F. A. & Pillai, R. S. Post-transcriptional gene silencing by siRNAs and miRNAs. *Curr. Opin. Struct. Biol.* **15**, 331–341 (2005).
6. Hutvagner, G. & Simard, M. J. Argonaute proteins: key players in RNA silencing. *Nature Rev. Mol. Cell Biol.* **9**, 22–32 (2008).
7. Kim, V. N., Han, J. & Siomi, M. C. Biogenesis of small RNAs in animals. *Nature Rev. Mol. Cell Biol.* **10**, 126–139 (2009).
8. Carthew, R. W. & Sontheimer, E. J. Origins and mechanisms of miRNAs and siRNAs. *Cell* **136**, 642–655 (2009).
9. Song, J. J., Smith, S. K., Hannon, G. J. & Joshua-Tor, L. Crystal structure of argonaute and its implications for RISC slicer activity. *Science* **305**, 1434–1437 (2004).
10. Parker, J. S., Roe, S. & Barford, D. Crystal structure of a PIWI protein suggests mechanisms for siRNA recognition and slicer activity. *EMBO J.* **23**, 4727–4737 (2004).
11. Yuan, Y. R. *et al.* Crystal structure of *A. aeolicus* argonaute, a site-specific DNA-guided endoribonuclease, provides insights into RISC-mediated mRNA cleavage. *Mol. Cell* **19**, 405–419 (2005).
12. Liu, J. *et al.* Argonaute2 is the catalytic engine of RNAi. *Science* **305**, 1437–1441 (2004).
13. Rivas, F. V. *et al.* Purified Ago2 and an siRNA form recombinant human RISC. *Nature Struct. Biol.* **12**, 340–349 (2005).
14. Parker, J. S., Roe, S. M. & Barford, D. Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex. *Nature* **434**, 663–666 (2005).
15. Ma, J.-B. *et al.* Structural basis for 5'-end-specific recognition of guide RNA by the *A. fulgidus* Piwi protein. *Nature* **434**, 666–670 (2005).
16. Ma, J.-B., Ye, K. & Patel, D. J. Structural basis for overhang-specific small interfering RNA recognition by the Paz domain. *Nature* **429**, 318–322 (2004).
17. Lingel, A., Simon, B., Izaurralde, E. & Sattler, M. Nucleic acid 3'-end recognition by the Argonaute2 Paz domain. *Nature Struct. Biol.* **11**, 576–577 (2004).
18. Martinez, J. & Tuschl, T. RISC is a 5'-phosphomonoester-producing RNA endonuclease. *Genes Dev.* **18**, 975–980 (2004).
19. Schwarz, D. S., Tomari, Y. & Zamore, P. D. The RNA-induced silencing complex is a Mg²⁺-dependent endonuclease. *Curr. Biol.* **14**, 787–791 (2004).
20. Elbashir, S. M., Lendeckel, W. & Tuschl, T. RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev.* **15**, 188–200 (2001).
21. Ameres, S. L., Martinez, J. & Schroeder, R. Molecular basis for target RNA recognition and cleavage by human RISC. *Cell* **130**, 101–112 (2007).
22. Rana, T. M. Illuminating the silence: understanding the structure and function of small RNAs. *Nature Rev. Mol. Cell Biol.* **8**, 23–36 (2007).
23. Parker, J. S. *et al.* Enhancement of the seed-target recognition step in RNA silencing by a PIWI/MID domain protein. *Mol. Cell* **33**, 204–214 (2009).
24. Parker, J. S. & Barford, D. Argonaute: a scaffold for the function of short regulatory RNAs. *Trends Biochem. Sci.* **31**, 622–630 (2006).
25. Patel, D. J. *et al.* Structural biology of RNA silencing and its functional implications. *Cold Spring Harb. Symp. Quant. Biol.* **71**, 81–93 (2006).
26. Tolia, N. H. & Joshua-Tor, L. Slicer and the argonautes. *Nature Chem. Biol.* **3**, 36–43 (2007).
27. Jinek, M. & Doudna, J. A three-dimensional view of the molecular machinery of RNA interference. *Nature* **457**, 405–412 (2009).
28. de Fougères, A., Vornlocher, H.-P., Maraganore, L. & Lieberman, J. Interfering with disease: a progress report on siRNA-based therapeutics. *Nature Rev. Drug Discov.* **6**, 443–453 (2007).
29. Castanotto, D. & Rossi, J. J. The promises and pitfalls of RNA-interference based therapeutics. *Nature* **457**, 426–433 (2009).
30. Wang, Y. *et al.* Structure of the guide-strand-containing argonaute silencing complex. *Nature* **456**, 209–213 (2008).
31. Wang, Y. *et al.* Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. *Nature* **456**, 921–926 (2008).
32. Mi, S. *et al.* Sorting of small RNAs into *Arabidopsis* argonaute complexes is directed by the 5'-terminal nucleotide. *Cell* **133**, 116–127 (2008).
33. Montgomery, T. A. *et al.* Specificity of ARGONAUTE7-miR390 interaction and dual functionality in TAS3 trans-acting siRNA formation. *Cell* **133**, 128–141 (2008).
34. Filipowicz, W. The nuts and bolts of the RISC machine. *Cell* **122**, 17–20 (2005).
35. Chi, S. W., Zang, J. B., Mele, A. & Darnell, R. B. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature* **460**, 479–486 (2009).
36. Nowotny, M., Gaidamakov, S. A., Crouch, R. J. & Yang, W. Crystal structures of RNase H bound to an RNA/DNA hybrid: Substrate specificity and metal-dependent catalysis. *Cell* **121**, 1005–1016 (2005).
37. Nowotny, M. Retroviral integrase superfamily: the structural perspective. *EMBO Rep.* **10**, 144–151 (2009).
38. Turner, D. H. Thermodynamics of base pairing. *Curr. Opin. Struct. Biol.* **6**, 299–304 (1996).
39. Freier, S. M. & Altmann, K. H. The ups and downs of nucleic acid duplex stability: structure-stability studies on chemically-modified DNA:RNA duplexes. *Nucleic Acids Res.* **25**, 4429–4443 (1997).
40. Verma, S. & Eckstein, F. Modified oligonucleotides: synthesis and strategy for users. *Annu. Rev. Biochem.* **67**, 99–134 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The research was supported by funds from the National Institutes of Health (NIH) and the Starr Foundation to D.J.P., and from the NIH, Starr Foundation and the Howard Hughes Medical Institute (HHMI) to T.T. We would like to thank the staff of NE-CAT beam lines at the Advanced Photon Source (APS), Argonne National Laboratory, and the X-29 beamline at the Brookhaven National Laboratory, supported by the US Department of Energy, for assistance with data collection. We thank Z. Wang for assistance with X-ray data collection at the APS.

Author Contributions Y.W. and G.S. expressed and purified wild-type *T. thermophilus* Ago and its catalytic mutants, and also grew crystals of the various ternary complexes. H.L. collected X-ray diffraction data on the various NE-CAT beam lines, and Y.W. solved the structures of these ternary complexes. D.J.P. supervised the structural studies. S.J. was responsible for the cleavage assays on Ago with modified DNA and RNA target strands, and G.S.W. purified the phosphorothioate diastereomers and quality controlled oligonucleotides, under the supervision of T.T. D.J.P. and T.T. were primarily responsible for writing the structural and biochemical contents of the paper, respectively, and all authors read and approved the submitted manuscript.

Author Information The structures of ternary complexes of *T. thermophilus* Ago have been submitted to the Protein Data Bank. The accession codes are: 3HO1 (mutant Ago(Asn 546)–12-nucleotide target RNA), 3HJF (mutant Ago(Glu 546)–15-nucleotide target RNA), 3HK2 (mutant Ago(Asn 478)–19-nucleotide target RNA), 3HM9 (wild-type Ago–19-nucleotide target RNA, 50 mM Mg), 3HVR (wild-type Ago–19-nucleotide target RNA, 80 mM Mg), and 3HXM (second crystal form of wild-type Ago–20-nucleotide target RNA containing two mismatches³¹). Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Correspondence and requests for materials should be addressed to D.J.P. (pateld@mskcc.org) or T.T. (ttuschl@mail.rockefeller.edu).

METHODS

Crystallization and data collection. Wild-type and mutant *T. thermophilus* Ago were prepared as described previously³⁰. Oligodeoxynucleotides were purchased from Invitrogen. RNA oligonucleotides were purchased from Dharmacon. For crystallization, *T. thermophilus* Ago was mixed with 5'-phosphorylated 21-nucleotide guide DNA at a 1:1.2 molar ratio, followed by the addition of different length target RNAs at a 1:1 molar ratio to the binary mixture, to form the ternary complex. All crystals were grown at 35 °C.

The mutant Ago protein complexes were crystallized by sitting-drop vapour diffusion method. Crystals of catalytic mutant Ago(Asn 546) complexed with 12-nucleotide target RNA were grown in a reservoir containing 2.5 mM spermine, 10 mM MgCl₂, 5 mM CaCl₂, 50 mM sodium cacodylate, pH 6.0, 10% (v/v) isopropanol. The crystals belong to space group C2, and there is one Ago complex in the asymmetric unit. Crystals of catalytic mutant Ago(Glu 546) complexed with 15-nucleotide target RNA were grown in a reservoir containing 1.3 M ammonium tartrate dibasic and 0.1 M Bis-Tris, pH 7.0. The crystals belong to space group P4₃2₁2, and there is one Ago complex in the asymmetric unit. Crystals of the catalytic mutant Ago(Asn 478) complexed with 19-nucleotide target RNA were obtained in a reservoir containing 1.0 M succinic acid, 0.1 M HEPES, pH 7.0, 1% (w/v) polyethylene glycol monomethyl ether 2,000. The crystals belong to space group P2₁2₁2₁, and there are two Ago complexes in the asymmetric unit.

Crystals of wild-type Ago complexed with 19-nucleotide target RNA with one bound divalent cation in the catalytic pocket were obtained with hanging-drop vapour diffusion method. The reservoir solution contained 50 mM MgCl₂, 1.0 M sodium tartrate, 50 mM Tris-HCl, pH 7.0. The crystals belong to space group P4₃2₁2, and there is one Ago complex in the asymmetric unit. With additional 30 mM MgCl₂ in both the reservoir and Ago protein, we obtained wild-type Ago complexed with 19-nucleotide target RNA and two bound divalent cations in the catalytic pocket. These crystals belong to space group P2₁2₁2₁, and there are two Ago complexes in the asymmetric unit.

Crystals of a second crystal form of wide-type Ago complexed with 20-nucleotide RNA target containing adjacent mismatches at the 10–11 step was grown under the same conditions as described previously³¹.

Diffraction data were collected on beamline NE-CAT ID-24C at the Advanced Photon Source (APS), Argonne National Laboratory and beamline X-29 at the Brookhaven National Laboratory. All data sets were integrated and scaled with the HKL2000 suite⁴¹ and data processing statistics are summarized in Supplementary Tables 1–3.

Structure determination and refinement. The structures of the complexes were solved by molecular replacement with the program PHASER⁴². The domains of the Ago 21-nucleotide guide DNA binary complex structure³⁰ without the linkers were used as search models. Model building was done using COOT⁴³, and refinement was done with CNS⁴⁴ and PHENIX⁴⁵. The final figures were created with Pymol (<http://pymol.sourceforge.net/>). The refinement statistics for all the Ago mutants and wild-type complexes are summarized in Supplementary Tables 1–3.

Oligonucleotides and separation of isomers. Phosphorothioate-modified and unmodified oligodeoxynucleotides were obtained from Integrated DNA technologies. R_p and S_p diastereomers were separated by HPLC using a Supelco Discovery C18 column (bonded phase silica 5 µm particle, 250 × 4.6 mm) following the general method described previously⁴⁶: buffer A was 0.1 M triethylammonium bicarbonate (TEAB, pH 7.5); buffer B was 40% acetonitrile in 0.1 M TEAB; flow rate was 1 ml min⁻¹. For the preparative scale ~20 optical density units (ODUs) (260 nm) of oligodeoxynucleotide (that is, 20 µl of a 1 mM stock solution) were loaded on the column. Diastereomers of positions 8'–9', 10'–11' and 11'–12' were separated using a two-step gradient, 0–20% B in 2 min followed

by 20–40% B in 40 min (0.5% change per min). The diastereomers of positions 9'–10' were more difficult to separate and the second step gradient was changed to 20–40% B in 80 min (0.25% change per min). Peak 1 was shown to be 97% pure by analytical HPLC (same conditions as preparative run, 0.3 ODU's injected). Peak 2 was shown to be 85% pure (Supplementary Fig. 26). Dithiothreitol (DTT) was added to the collected peak fractions (1 µl 100 mM DTT to about 2 ml fraction) before dry down to minimize oxidation of the phosphorothioate. Co-evaporation with methanol was repeated three times to remove residual TEAB buffer. The dried-down material was resuspended in 50 µl water and ethanol precipitated to remove DTT. In each case peak 1 is the R_p form and peak 2 is the S_p form, consistent with ref. 47. The identity of the purified diastereomers was confirmed by snake venom phosphodiesterase/alkaline phosphatase treatment and subsequent HPLC; the S_p-configured dinucleotide was more resistant to phosphodiesterase compared to the R_p-configured dinucleotide⁴⁸ (Supplementary Fig. 27). HPLC for separation of nucleosides and dinucleotide phosphorothioates used buffer A as 0.1 M triethylammonium acetate in 3% acetonitrile, and buffer B as 90% aqueous acetonitrile. The elution was performed using a stepwise gradient starting at 0% B for 15.5 min, followed by 19.5 min of 10% B and 30 min of 100% B at a flow rate of 0.5 ml min⁻¹, using the same HPLC column as indicated earlier. The first four peaks are digested monomers, consistent for each oligodeoxynucleotide. Later peaks (after 30 min) are undigested dimers along with some baseline noise. Elution times of these later peaks depended on the dimer sequence and phosphorothioate configuration. In each case the R_p form was more digested than the S_p form (Supplementary Fig. 27).

Cleavage activity assay of *T. thermophilus* Ago. Recombinant *T. thermophilus* Ago (0.5 µM final concentration) was incubated with a reaction mixture containing 10 mM HEPES-KOH, pH 7.5, 100 mM NaCl, 0.5 µM guide strand, and 5 mM of CaCl₂, MgCl₂ or MnCl₂ for 30 min at 55 °C in a volume of 10 µl. Then, 5'-³²P-labelled DNA target was added to obtain the indicated final concentrations. For single turnover conditions (0.5 µM target strand) or multiple turnover conditions (5 µM target strand), unlabelled DNA target was spiked with radioactive target at a concentration of approximately 0.01 µM. The incubation was continued at 75 °C in a total volume of 15 µl. The reaction was stopped by the addition of 15 µl Stop solution (95% formamide, 50 mM EDTA and 0.02% bromophenol blue). The cleavage products were resolved on a 12% denaturing polyacrylamide gel, and radioactivity was monitored by phosphorimaging.

- Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Meth. Enzymol.* **276**, 307–326 (1997).
- McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
- Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
- Brünger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
- Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58**, 1948–1954 (2002).
- Tonelli, M. *et al.* Dynamic NMR structures of [R_p]- and [S_p]-phosphorothioated DNA-RNA hybrids: is flexibility required for RNase H recognition? *Biophys. J.* **85**, 2525–2538 (2003).
- Thorogood, H., Grasby, J. A. & Connolly, B. A. Influence of the phosphate backbone on the recognition and hydrolysis of DNA by the EcoRV restriction endonuclease. A study using oligodeoxynucleotide phosphorothioates. *J. Biol. Chem.* **271**, 8855–8862 (1996).
- Burgers, P. M. & Eckstein, F. Absolute configuration of the diastereomers of adenosine 5'-O-(1-thiotriphosphate): consequences for the stereochemistry of polymerization by DNA-dependent RNA polymerase from *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **75**, 4798–4800 (1978).

ARTICLES

Role of the polycomb protein EED in the propagation of repressive histone marks

Raphael Margueron^{1*}, Neil Justin^{2*}, Katsuhito Ohno^{3*}, Miriam L. Sharpe^{2*}, Jinsook Son¹, William J. Drury III¹, Philipp Voigt¹, Stephen R. Martin², William R. Taylor², Valeria De Marco², Vincenzo Pirrotta³, Danny Reinberg¹ & Steven J. Gamblin²

Polycomb group proteins have an essential role in the epigenetic maintenance of repressive chromatin states. The gene-silencing activity of the Polycomb repressive complex 2 (PRC2) depends on its ability to trimethylate lysine 27 of histone H3 (H3K27) by the catalytic SET domain of the EZH2 subunit, and at least two other subunits of the complex: SUZ12 and EED. Here we show that the carboxy-terminal domain of EED specifically binds to histone tails carrying trimethyl-lysine residues associated with repressive chromatin marks, and that this leads to the allosteric activation of the methyltransferase activity of PRC2. Mutations in EED that prevent it from recognizing repressive trimethyl-lysine marks abolish the activation of PRC2 *in vitro* and, in *Drosophila*, reduce global methylation and disrupt development. These findings suggest a model for the propagation of the H3K27me3 mark that accounts for the maintenance of repressive chromatin domains and for the transmission of a histone modification from mother to daughter cells.

The fate of a cell is specified by its gene expression profile, often set early in development and maintained throughout the lifetime of the cell by epigenetic mechanisms. The polycomb group of proteins functions by silencing inappropriate expression by maintaining a repressive epigenetic state¹. It is thought that the PRC2-mediated trimethylation of lysine 27 on histone H3 (H3K27me3) has a crucial role in marking repressive chromatin domains, whereas PRC1 is important for effecting transcriptional repression. Thus, once established, H3K27 trimethylation is the epigenetic mark for maintaining transcriptional repression. Mechanisms are therefore required to maintain this mark in repressed chromatin domains in non-dividing cells and to restore it after the twofold dilution caused by DNA replication in dividing cells. However, it is not yet clear how PRC2 complexes recognize previously marked sites and how they accurately propagate these repressive marks to unmodified nucleosomes deposited during DNA replication.

The histone lysine methyltransferase (HKMT) activity of the PRC2 complex resides in the SET-domain-containing protein EZH2 (refs 2–5), but activity requires the other subunits of the core complex; the zinc-finger-containing SUZ12 and the WD40 repeat proteins EED and RbAp48 (also known as CAF1). In certain contexts, the PHD-domain-containing protein PHF1 plays an important part in modulating the HKMT activity of PRC2 (refs 6, 7). In this work we have examined the structure and biochemistry of EED, and determined the role of its homologue ESC in *Drosophila* development. From this we have established that the EED subunit of PRC2 binds to repressive methyl-lysine marks, ensuring the propagation of H3K27 trimethylation on nucleosomes by allosterically activating the methyltransferase activity of the complex (see Supplementary Fig. 1).

The aromatic cage of EED binds repressive chromatin marks

We crystallized a truncated version of EED (residues 77 to 441, hereafter Δ EED) and used selenomethionine-substituted Δ EED to solve the

structure. The WD40-repeats of Δ EED fold into a seven-bladed β -propeller domain with a central pocket on either end (Fig. 1), as seen previously⁸. We noticed unaccounted electron density in one of these pockets; our crystallization mixture included a non-detergent sulphobetaine additive, NDSB-195, which we were able to build into the extra electron density. Because the quarternary amine of the sulphobetaine resembled a trimethylated lysine side chain⁹ we reasoned that EED might bind to trimethylated lysine residues on the N-terminal tails of histones.

Histone lysine residues methylated *in vivo* include H3K4, H3K9, H3K27, H3K36, H3K79, H4K20 and H1K26. We measured the binding

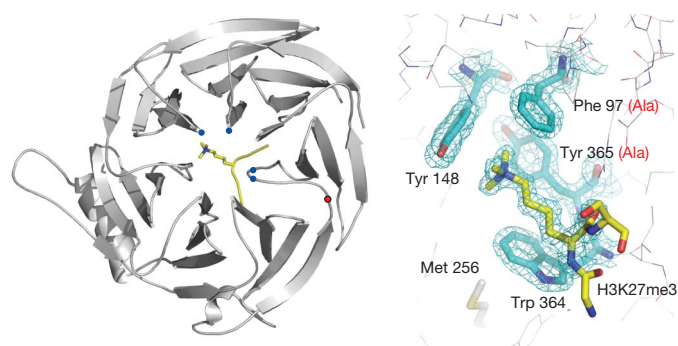


Figure 1 | Trimethyl-lysine binding to an aromatic cage on EED. Ribbons representation of the EED–H3K27me3 complex, in which EED is in grey and the histone peptide is in yellow with its methyl-lysine side chain shown in stick representation. The C α positions of the aromatic cage are shown as blue circles, and the C α position of Tyr 358 by a red circle. The right panel shows the methyl-lysine-binding site with $2F_o - F_c$ electron density for the four cage residues and the H3K27me3 peptide. Designed mutations to the cage are shown in red in parentheses. The side chain of Met 256 is also shown; this is equivalent to Met 236 in ESC, which has been identified from classical genetic screens in *Drosophila* as essential for the function of EED.

¹Howard Hughes Medical Institute and Department of Biochemistry, New York University Medical School, 522 First Avenue, New York, New York 10016, USA. ²MRC National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK. ³Department of Molecular Biology and Biochemistry, Rutgers University, Nelson Laboratories, 604 Allison Road, Piscataway, New Jersey 08854, USA.

*These authors contributed equally to this work.

affinity of Δ EED to trimethylated versions of these lysine residues using synthetic peptides by fluorescence competition assays. Δ EED bound to H1K26me₃, H3K9me₃, H3K27me₃ and H4K20me₃ peptides with dissociation constant (K_d) values ranging from 10 to 45 μ M, and the binding became approximately fourfold weaker for each successive loss of a methyl group from the methyl-lysine (Supplementary Table 1). Notably, Δ EED did not bind appreciably to H3K4me₃, H3K36me₃ or H3K79me₃ 'marks' associated with active transcription¹⁰. We validated these results by isothermal titration calorimetry (Supplementary Table 1 and Supplementary Fig. 2b), and there is good agreement between the two independent methods.

Next, we solved the structure of Δ EED co-crystallized with H1K26me₃, H3K9me₃, H3K27me₃ and H4K20me₃ peptides (Supplementary Table 2 and Supplementary Fig. 3). The peptides in the four co-crystal structures adopt similar, largely extended structures and all exploit the aromatic cage of Δ EED to recognize the trimethyl-lysine residue (Fig. 1 and Supplementary Fig. 4). This is the first example of such a binding site on a β -propeller domain and it consists of three aromatic side-chains, Phe 97, Tyr 148 and Tyr 365 (Fig. 1). The trimethyl-ammonium group of the lysine is inserted into this cage and is stabilized by van der Waals and cation- π interactions. A fourth aromatic side-chain (Trp 364) interacts with the aliphatic moiety of the lysine side chain by hydrophobic interactions (Figs 1, 2 and Supplementary Fig. 5). Adjacent to the methyl-lysine pocket, Δ EED makes two hydrogen-bond interactions with carbonyls on the peptides (Fig. 2a). First, the main-chain carbonyl of the methyl-lysine residue forms hydrogen bonds with the side chain of Arg 414. Second, the main-chain carbonyl of the residue immediately amino-terminal to it, makes a hydrogen bond with the main-chain amide of Trp 364. The residues flanking the methyl-lysine

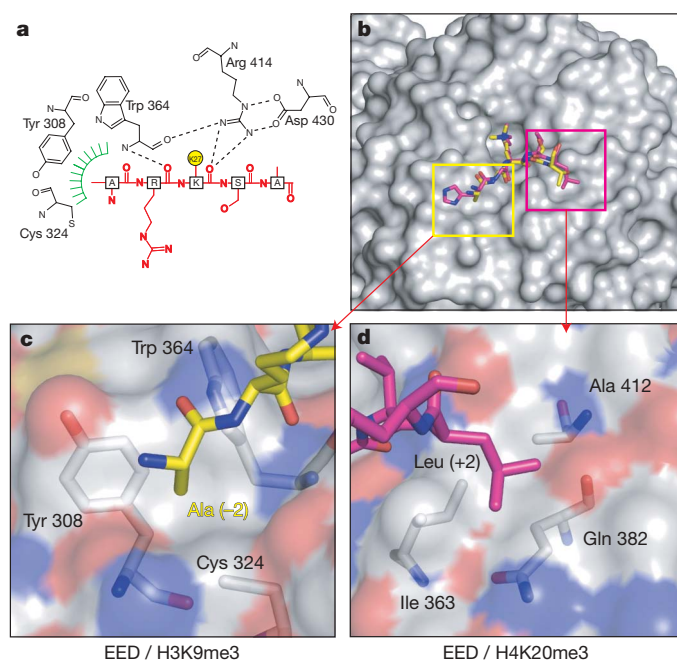


Figure 2 | Interactions between EED and trimethylated histone peptides. **a**, Schematic representation of the interaction between EED and H3K27me₃. For clarity, the aromatic methyl-lysine-binding cage has been omitted and the methylated lysine side-chain shown as a yellow circle. Hydrogen bonds from the main-chain carbonyl of the methyl-lysine with EED, and the residue immediately N-terminal to it, are shown as dashed lines. The green hatching indicates the hydrophobic pocket on EED that accommodates the alanine side chain two residues before the methyl-lysine. **b**, EED is shown in surface representation with a composite of two of its cognate peptides shown in stick representation and coloured yellow (H3K9me₃) and pink (H4K20me₃). **c**, The pocket on EED that accommodates Ala (-2) from the H3K9me₃ peptide is shown. **d**, The other pocket that contains a Leu (+2) from the H4K20me₃ peptide. The EED surface is coloured according to atom type.

residue, at the -1 and +1 positions, are oriented away from the protein, whereas the next residues, at the -2 and +2 positions, make important contacts (Fig. 2). Comparison of the four complexes (Fig. 2 and Supplementary Figs 2A and 4) suggests an important role for two distinct hydrophobic interaction sites (Fig. 2b). H1K26, H3K9 and H3K27 each have an alanine residue two amino acids N-terminal to the lysine (-2), which fits into a small pocket on the surface of EED formed by the hydrophobic moieties of Trp 364, Tyr 308 and Cys 324 (Fig. 2c). The size of this pocket is sufficient to accommodate an alanine residue but not larger hydrophobic residues. In the case of H4K20 peptide—the only one of the four that bound to Δ EED and lacks an alanine at -2—its binding is facilitated by an alternative hydrophobic interaction between the leucine residue in the +2 position of the peptide with a second hydrophobic pocket formed by residues Ile 363, Ala 412 and the γ -carbon of Gln 382 of EED (Fig. 2d). It seems that the ability to exploit one of these two small hydrophobic pockets is an important component of the specificity of EED towards the methyl-lysine marks associated with repressive chromatin. However, the affinity of EED for these modified peptides is relatively modest, and it is likely that this interaction only becomes physiologically relevant in association with the histone-binding activity of other components of the PRC2 complex, as suggested by earlier work on *Drosophila* PRC2 (ref. 11).

To probe the physiological role of the aromatic cage of EED, site-directed mutants of several of the cage residues were created. Mutations of Phe 97, Trp 364 and Tyr 365 to alanine produced well-behaved protein, and competition experiments showed that the Trp364Ala and Tyr365Ala mutations had no detectable binding to H1K26me₃ peptides, whereas Δ EED Phe97Ala bound about eightfold more weakly than wild-type Δ EED to histone peptides (Supplementary Table 1). As a control for the effect of mutation of an aromatic residue on the EED structure that is not involved in the aromatic cage, we also generated the mutation Tyr358Ala (Fig. 1); binding by this mutant was reduced by about twofold (Supplementary Table 1).

PRC2-EZH2 and EED specifically bind H3K27me₃ nucleosomes

Next, we used nucleosome arrays reconstituted with chemically modified histones that carry a single modification of the four possible methylation states of H3K27, H3K36 or H3K9 (ref. 12). The nucleosome arrays were incubated with full-length His-tagged EED protein followed by nickel-nitrilotriacetic acid (Ni-NTA) pull-down assays. Western blotting for H3 and EED demonstrated an interaction between EED and nucleosomes containing H3K27me₃ (Fig. 3). This interaction was specific as EED was not able to interact with chromatin reconstituted with histones containing the different levels of H3K36 methylation (Supplementary Fig. 8), but did bind to chromatin trimethylated on H3K9 (data not shown). Interestingly, the truncated Δ EED protein tested in the peptide-binding experiments also failed to

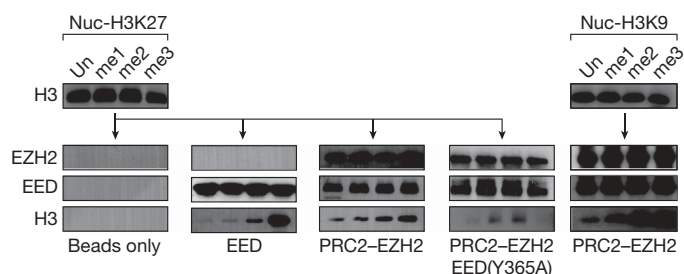


Figure 3 | EED and PRC2 interaction with chromatin. Pull-down experiment to analyse the interaction between EED, PRC2-EZH2 wild type or reconstituted with EED(Tyr365Ala) and H3K27-modified chromatin (left), or between PRC2-EZH2 wild type and H3K9-modified chromatin (right). Nuc, nucleosomes; un, unmethylated. Note that the 'beads only' control for interaction with H3K9-modified chromatin is not shown but was identical to the control shown in the figure.

interact with nucleosomes (Supplementary Fig. 8). Presumably, the diminished binding is due to the absence of a previously characterized H3-binding site within the N terminus of EED¹³, which may act together with the methyl-lysine-binding site to achieve stable binding. Given that other subunits of PRC2 contact histones and thus modulate chromatin binding, we repeated the nucleosome-binding experiment using a PRC2 complex purified from insect cells co-infected with baculovirus expressing each of the subunits (Fig. 4a, left). Although, as expected, the reconstituted PRC2 complex showed some binding to unmodified chromatin, the complex bound considerably tighter to chromatin carrying the H3K27me3 or the H3K9me3 modification (Fig. 3). Interestingly, PRC2 reconstituted with EED carrying the Phe97Ala or Tyr365Ala substitution does not show binding to chromatin under these conditions, with either methylated or unmodified nucleosomes (Fig. 3 and data not shown). Together, our results demonstrate that the aromatic cage in EED is critical for the PRC2 complex to bind to repressive marks, through its specific recognition of defined (repressive) trimethylated-lysine residues.

Trimethylated repressive marks stimulate PRC2 activity

Because a probable function for the binding of PRC2 to trimethylated lysine would be to contribute to the propagation of the H3K27me3

mark, we performed HKMT assays using recombinant oligonucleosomes in the presence of methylated peptides. The addition of unmodified or monomethylated H3K27 peptides did not significantly affect the enzymatic activity of PRC2, but trimethylated peptides activated it by about sevenfold (Fig. 4a). Stimulation of enzymatic activity by the H3K27me3 peptide reached a plateau around 100 μ M, and half-maximum stimulation is achieved at 30–40 μ M (Supplementary Fig. 7), which is in good agreement with the dissociation constant determined for Δ EED and the H3K27me3 peptide (Supplementary Table 1) and gives us strong confidence that the binding event we observe with purified, truncated EED is closely correlated with the allosteric activation mechanism. We also determined the Michaelis parameters for PRC2 in the presence of variously methylated H3K27 peptides (Fig. 4b, c). During titrations of S-adenosyl-methionine (SAM) we observed a marked increase in the maximum reaction rate (V_m) in the presence of the H3K27me3 peptide. A similar result was observed with titration of nucleosomes (Fig. 4c). Notably, in both cases the substrate concentration required to achieve the half-maximal reaction rate (K_m) is not significantly affected by the incubation with peptides.

To ascertain whether the observed stimulation was EED-mediated, mutant PRC2 complexes containing EED(Phe97Ala) or EED

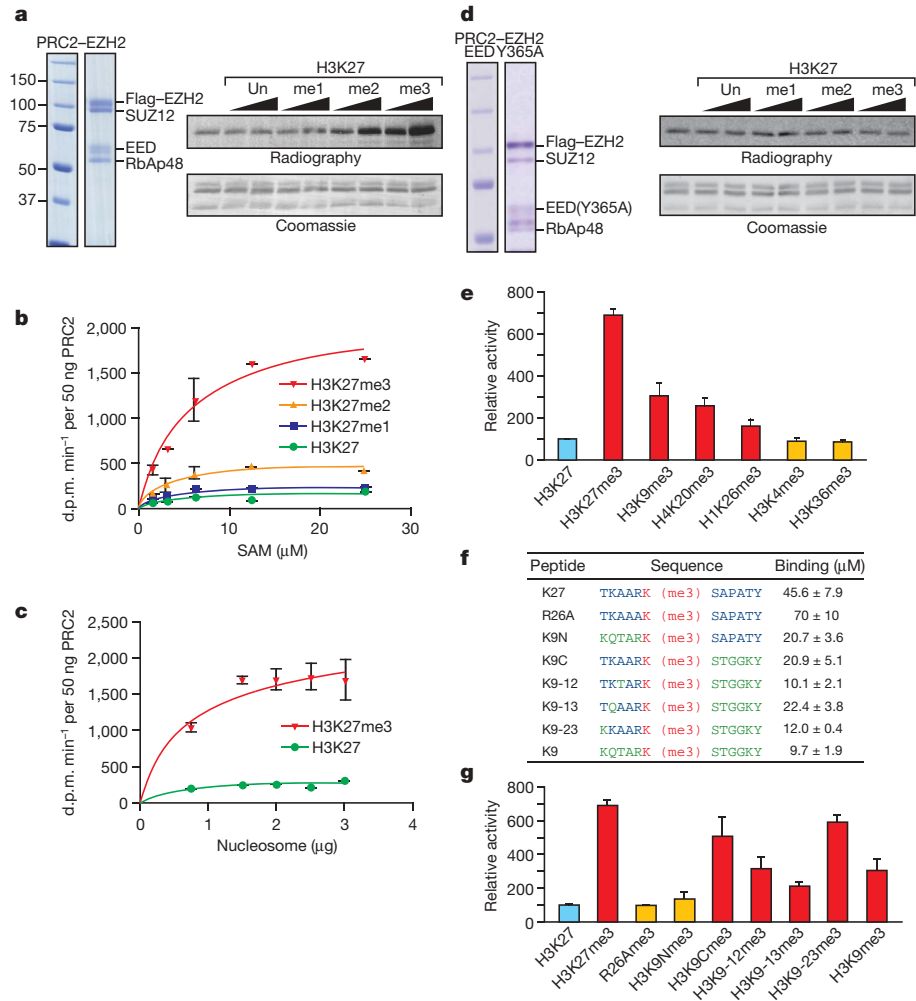


Figure 4 | Peptide mimicking repressive marks stimulates PRC2 activity. **a**, Left, Coomassie blue staining of reconstituted PRC2-EZH2 complex. Right, HKMT assay with PRC2-EZH2 alone or in the presence of 10 and 40 μ M H3K27, unmodified, mono, di or trimethylated peptides. **b**, Titration of the methyl donor (S-adenosyl-methionine) in the presence of H3K27me0/1/2/3 peptides. d.p.m., disintegration per min. **c**, Nucleosome titration in the presence of H3K27me0/3 peptides. **d**, Left, coomassie blue staining of

reconstituted PRC2-EZH2 EED(Tyr365Ala) complex. Right, HKMT assay with the corresponding complex in the same condition as **a**. **e**, Relative PRC2 histone methyltransferase activity in the presence of various peptides as indicated. **f**, Table indicating the peptides used for the stimulation study as well as their K_d values (μ M) for Δ EED binding. **g**, Relative PRC2 histone methyltransferase activity in the presence of various peptides as indicated.

(Tyr365Ala) were reconstituted (Fig. 4d and Supplementary Fig. 8). These mutant recombinant PRC2 complexes retain a similar basal activity to wild-type, but neither mutant recombinant PRC2 was stimulated by the addition of H3K27me3 peptides (Fig. 4d). Our data also show that the H3K9me3, H4K20me3 and H1K26me3 peptides were all able to stimulate PRC2 activity to some extent, whereas the H3K4me3 and H3K36me3 peptides were ineffectual (Fig. 4e). However, we noticed that the binding affinity to EED and stimulation of PRC2 activity do not strictly correlate (that is, H3K9me3 has a good binding affinity for EED but stimulates PRC2 activity relatively poorly). To investigate the role of histone sequence in binding/activation we first mutated the arginine residue at the -1 position—present in all four histone peptides that activate the methyltransferase activity of PRC2—to alanine (Arg26Ala of H3K27me3 in Fig. 4f, g). Remarkably, although the binding of this mutant peptide to ΔEED is only reduced by about 1.5-fold (Fig. 4f), it is no longer able to activate PRC2 HKMT activity (Fig. 4g), demonstrating that repressive-histone-peptide binding to the aromatic cage of EED is necessary, but not

sufficient for PRC2 activation. To further test this model we made a series of chimaeric and mutant peptides that show that the lysine at -4, the alanine at -3 and the arginine at -1 are not important for binding to EED but are key to the activation of PRC2. We propose that these are the residues that mediate an interaction with another part of the PRC2 complex that leads to its activation.

PRC2 function requires recognition of repressive marks

To evaluate the importance of EED binding to trimethylated marks *in vivo*, we turned to ESC, the EED homologue in *Drosophila*, and tested the effect of mutating its aromatic cage. We reconstituted the *Drosophila* PRC2 complex and showed that addition of H3K27me2/3 peptides to the HKMT assay resulted in a robust stimulation of PRC2 enzymatic activity (Fig. 5a). ESC is required throughout development, but in the early embryo the maternal stock of *esc* product is critical, as evidenced by the resultant derepression of homeotic genes in embryos produced by *esc*⁻ mothers^{14,15}. At later stages of development, PRC2 activity is sustained through the overlapping participation of ESC and

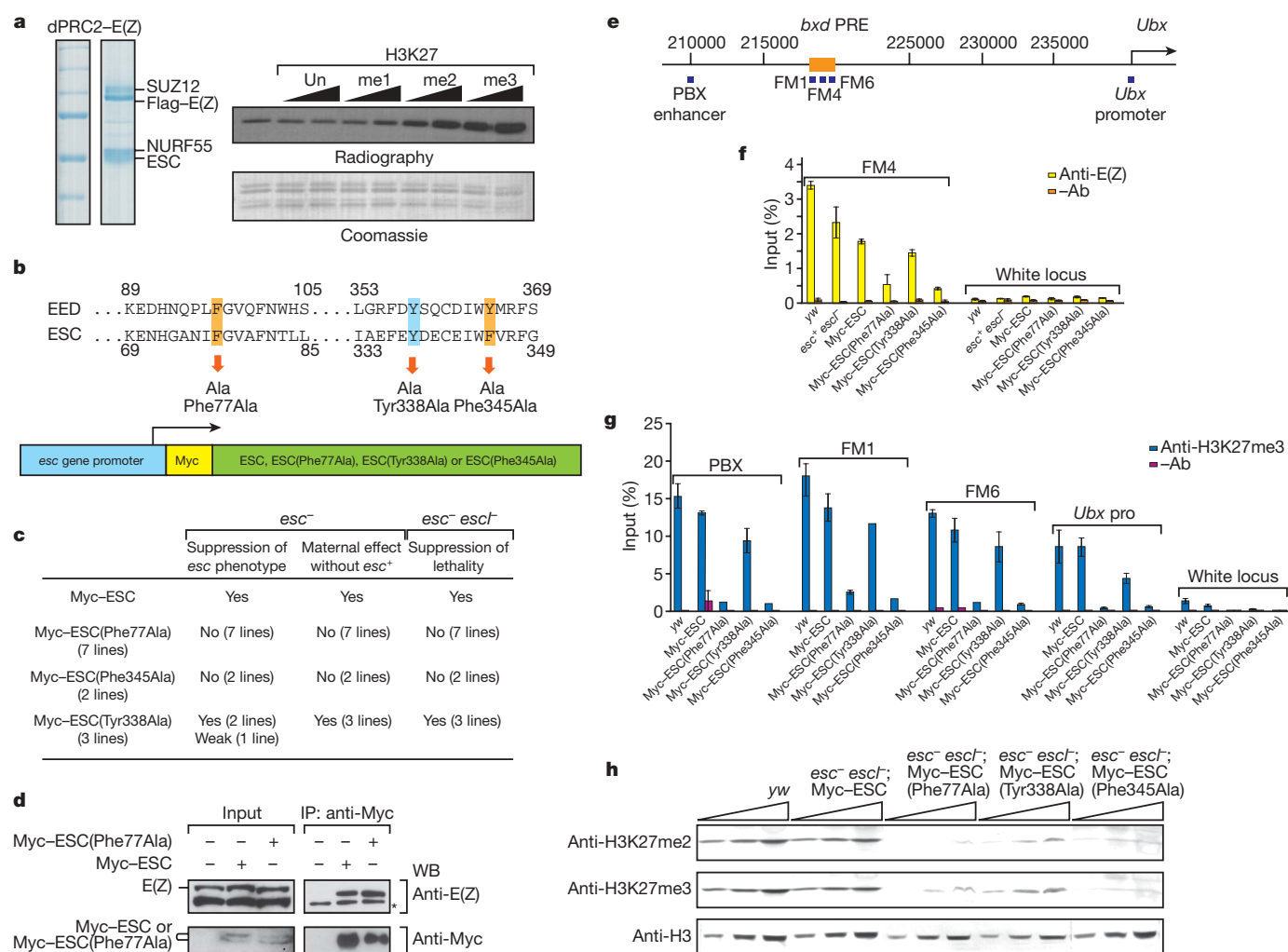


Figure 5 | The aromatic cage in *Drosophila* ESC is important for its *in vivo* function. **a**, Left, Coomassie blue staining of reconstituted dPRC2-E(Z) complex. Right, HKMT assay with dPRC2-E(Z) and H3K27me0/1/2/3 peptides. **b**, Top, amino acid residues Phe 77, Tyr 338 and Phe 345 that were mutated to Ala in *Drosophila* ESC and the corresponding residues in EED. Bottom, schematic representation of transposon constructs. **c**, Rescue experiment. See Supplementary Figs 9 and 10 for crossing scheme details. Several independent lines were examined for each transgenic construct and showed the same phenotype except for one line of Myc-ESC(Tyr338Ala). In the case of Myc-ESC(Phe345Ala), transgenes were inserted at φC31 att sites at 68E and 86Fb, respectively. For direct comparison, wild-type Myc-ESC lines were also established at the same chromosomal location and showed

the same results as lines established by conventional P-element transformation. **d**, Immunoprecipitation using ovarian extracts. ESC aromatic cage mutation does not impair binding to E(Z). The double Myc-ESC bands are caused by phosphorylation of ESC. **e**, Scheme indicating the genomic location of primers used for ChIP. **f**, ChIP analysis of E(Z) binding to the *bxd* PRE (FM4) in homozygous *esc*⁶ *esc*^{d01514} expressing wild-type or aromatic cage mutant Myc-ESCs. *yw* indicates the wild-type stock with endogenous wild-type ESC and ESCL. **g**, ChIP analysis of the H3K27me3 distribution at four sites in the *Ubx* gene. **h**, Histone H3K27 methylation in *esc*⁶ *esc*^{d01514} double-mutant larvae expressing wild-type or aromatic cage mutant Myc-ESCs.

its close homologue ESCL^{15,16}. Overexpression of ESC in the ovaries (for example, in a female with one extra *esc* copy) can supply enough function to allow the development of *esc* embryos, producing flies that are virtually normal except for the eponymous extra sex combs in males. We constructed mutations affecting the aromatic cage: Phe77Ala (equivalent to Phe97 in EED) and Phe345Ala (equivalent to Tyr365 in EED), as well as Tyr338Ala (equivalent to Tyr358Ala in EED) just preceding the aromatic cage (Fig. 5b), and expressed Myc-tagged wild-type or mutant *esc* transgenes under the control of the *esc* promoter (Fig. 5b). Although the wild-type transgene rescued the extra sex comb phenotype almost completely (217 out of 218 males counted), the aromatic cage mutant transgenes were ineffectual (no rescue in several hundred males examined) (Fig. 5c). Flies lacking both zygotic ESC and ESCL in these crosses produce larvae with poorly developed brain and imaginal discs, which die when they pupate. This lethality is completely rescued by one copy of the wild-type *esc*>Myc-ESC transgene. In contrast, none of the aromatic cage mutant transgenes were able to rescue the lethality even when present in two copies (Fig. 5c), although the *esc*>Myc-ESC(Phe77Ala) transgene alleviated the brain and imaginal disc phenotypes (data not shown). Of note, zygotic expression of the Phe345Ala transgene impaired the contribution of wild-type *esc* indicating that this mutant acts as a dominant negative. The failure of the mutant Myc-ESC to rescue is not due to instability or the inability to be incorporated into a PRC2 complex: the ESC mutants were expressed at levels comparable to that of the wild type. Furthermore, immunoprecipitation experiments showed that the mutant ESCs co-immunoprecipitated with endogenous E(Z) as efficiently as the wild-type protein (Fig. 5d). To determine whether the mutant ESCs affected PRC2 function with respect to its gene targeting or activity, we performed chromatin immunoprecipitation (ChIP) followed by quantitative PCR with the primer sets indicated in Fig. 5e. Immunoprecipitation using anti-E(Z) shows that wild-type Myc-ESC is nearly as effective as endogenous ESC (compare with the *esc*⁺ *escl*⁻ chromatin), whereas PRC2 complex with Myc-ESC-bearing mutations in the aromatic pocket is recruited less efficiently to the *Ubx* polycomb response element (PRE) (Fig. 5f). Chromatin immunoprecipitation with anti-H3K27me3 antibodies also shows that wild-type Myc-ESC is nearly as effective as endogenous ESC (*yw*) in trimethylating H3K27 in the *Ubx* upstream enhancer region (PBX, -30 kilobases (kb)), in the vicinity of the PRE (FM1, FM6, -23 kb) or at the *Ubx* promoter. Notably, the mutant ESCs are deficient in the extent of H3K27me3 (Fig. 5g), and this decrease correlates with the phenotypes described in Fig. 5c. Importantly, the observed effects are due to the aromatic cage, as a mutation of Tyr338Ala, which is not important for cage formation, had no effect (Fig. 5g). Finally, we analysed the global levels of H3K27 methylation by western blot (Fig. 5h). We observed an almost complete loss of H3K27me3 in extracts from *esc*⁻ *escl*⁻ larvae expressing the mutant ESCs. Perhaps surprisingly, the H3K27me2 levels were equally strongly affected.

Discussion

Chromatin domains are distinguished by the presence of a characteristic set of marks. When these marks are used to sustain an epigenetic state, eukaryotic cells must have the means of propagating these marks through cellular division and of ensuring that they obey appropriate boundaries during development. That PRC2 might recognize the chromatin mark it sets was anticipated by Hansen *et al.*¹⁷ who reported that PRC2 binds to H3K27me3, although this study did not address the mechanism for the propagation of H3K27me3. Our work shows the structural and functional basis for epigenetic self-renewal, and leads us to conclude that PRC2 readout of H3K27me3 (and to a lesser extent other 'repressive' marks) is key to the propagation of this repressive mark.

A combination of aromatic and hydrophobic residues is commonly used by proteins that recognize methylated lysine residues and has been found in chromo-, tudor- and plant homeo-domains (PHD)^{18–20}, but

no such arrangement has previously been described for any WD40-repeat-containing protein (for example, Supplementary Fig. 7 and ref. 21). Sequence analysis across the family of β -propeller domains leads us to conclude that the ability of EED to specifically recognize repressive methyl-lysine marks is a feature, limited among WD40 proteins, to EED-related molecules (Supplementary Fig. 6).

This methyl-lysine interaction provides an extra contribution to nucleosome binding that is mainly driven by a combination of contacts from other subunits of PRC2; RbAp48 binds to histone H4 (refs 22, 23) and the N-terminal domain of EED binds to H3 (ref. 13), and it may well be that these different interactions act cooperatively. In *Drosophila*, recruitment of PRC2 may also be facilitated by certain DNA-binding factors^{24,25}. Our *Drosophila* experiments show that when the *Drosophila* EED orthologue ESC bears mutations in the aromatic cage, the recruitment of PRC2 to the PRE is less effective, as shown by the drop in E(Z) binding to the *bxd* PRE, the massive reduction in the global level of H3K27me2/3 and by the phenotype of the Phe77Ala and Phe345Ala mutants. Our chromatin modification assays suggest that a major effect of EED binding to repressive methyl-lysine marks is the stimulation of PRC2 methyltransferase activity, thus providing a mechanism for the propagation of this mark. Thus, when PRC2 is recruited to appropriate chromatin domains, the presence of pre-existing H3K27me3 marks on neighbouring nucleosomes activates the complex to carry out further methylation of unmodified H3K27 (Supplementary Fig. 1). Accordingly, a polycomb group target gene that had been repressed in one cell cycle will tend to be repressed again in the next cell cycle, and previously active genes will be left unmodified at H3K27. We propose that the ability to recognize a previously established mark that triggers its renewal is a feature that will be found in other epigenetic mechanisms mediated by histone modifications.

METHODS SUMMARY

N-terminally truncated EED (Δ EED) was expressed as a glutathione S-transferase (GST) fusion protein in *Escherichia coli*. Crystals were grown by the hanging drop method using 4.0 M sodium formate as a precipitant, together with NDSB-195 or as a complex with histone peptides. Diffraction data were processed using Denzo and Scalepack, the native structure was solved by single wavelength anomalous dispersion (SAD) and built by ARP/wARP. Histone complexes were solved by molecular replacement and refined using Refmac5 with manual model building using O or Coot. The affinity of wild-type EED for histone peptides was determined by competition fluorescence spectroscopy or isothermal titration calorimetry (ITC). Fluorescence spectroscopy was performed at 20 °C using a SPEX FluoroMax fluorimeter; dansyl-labelled peptides bound to EED were competed off by excess cold peptide and the change in fluorescence was monitored. ITC reactions were performed at 20 °C and used to verify EED wild-type binding and also to determine the affinity of EED mutants to trimethylated histone peptides. Recombinant PRC2 complexes were purified from insect cells after infection with baculovirus. Chromatin for interaction and HKMT assay was refolded by salt dialysis. Histone H3 carrying the Cys110Ala mutation was chemically modified using the method described previously¹². ESC transgene construction, *esc* and *escl* mutant fly lines have been described previously¹⁵. Chromatin immunoprecipitation and analysis by quantitative PCR were done according to refs 26 and 27.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 23 June; accepted 13 August 2009.

Published online 20 September 2009.

- Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B. & Cavalli, G. Genome regulation by polycomb and trithorax proteins. *Cell* **128**, 735–745 (2007).
- Czermin, B. *et al.* *Drosophila* enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. *Cell* **111**, 185–196 (2002).
- Cao, R. *et al.* Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298**, 1039–1043 (2002).
- Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P. & Reinberg, D. Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes Dev.* **16**, 2893–2905 (2002).
- Müller, J. *et al.* Histone methyltransferase activity of a *Drosophila* Polycomb group repressor complex. *Cell* **111**, 197–208 (2002).

6. Nekrasov, M. *et al.* Pcl-PRC2 is needed to generate high levels of H3-K27 trimethylation at Polycomb target genes. *EMBO J.* **26**, 4078–4088 (2007).
7. Sarma, K., Margueron, R., Ivanov, A., Pirrotta, V. & Reinberg, D. Ezh2 requires PHF1 to efficiently catalyze H3 lysine 27 trimethylation *in vivo*. *Mol. Cell. Biol.* **28**, 2718–2731 (2008).
8. Han, Z. *et al.* Structural basis of EZH2 recognition by EED. *Structure* **15**, 1306–1315 (2007).
9. Schiefner, A. *et al.* Cation- π interactions as determinants for binding of the compatible solutes glycine betaine and proline betaine by the periplasmic ligand-binding protein ProX from *Escherichia coli*. *J. Biol. Chem.* **279**, 5588–5596 (2004).
10. Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693–705 (2007).
11. Nekrasov, M., Wild, B. & Muller, J. Nucleosome binding and histone methyltransferase activity of *Drosophila* PRC2. *EMBO Rep.* **6**, 348–353 (2005).
12. Simon, M. D. *et al.* The site-specific installation of methyl-lysine analogs into recombinant histones. *Cell* **128**, 1003–1012 (2007).
13. Tie, F., Stratton, C. A., Kurzahls, R. L. & Harte, P. J. The N terminus of *Drosophila* ESC binds directly to histone H3 and is required for E(Z)-dependent trimethylation of H3 lysine 27. *Mol. Cell. Biol.* **27**, 2014–2026 (2007).
14. Struhl, G. & Brower, D. Early role of the *esc*⁺ gene product in the determination of segments in *Drosophila*. *Cell* **31**, 285–292 (1982).
15. Ohno, K., McCabe, D., Czermin, B., Imhof, A. & Pirrotta, V. ESC, ESCL and their roles in Polycomb Group mechanisms. *Mech. Dev.* **125**, 527–541 (2008).
16. Kurzahls, R. L., Tie, F., Stratton, C. A. & Harte, P. J. *Drosophila* ESC-like can substitute for ESC and becomes required for Polycomb silencing if ESC is absent. *Dev. Biol.* **313**, 293–306 (2008).
17. Hansen, K. H. *et al.* A model for transmission of the H3K27me3 epigenetic mark. *Nature Cell Biol.* **10**, 1291–1300 (2008).
18. Huang, Y., Fang, J., Bedford, M. T., Zhang, Y. & Xu, R.-M. Recognition of histone H3 lysine-4 methylation by the double tudor domain of JMJD2A. *Science* **312**, 748–751 (2006).
19. Li, H. *et al.* Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. *Nature* **442**, 91–95 (2006).
20. Peña, P. V. *et al.* Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. *Nature* **442**, 100–103 (2006).
21. Southall, S. M., Wong, P. S., Odho, Z., Roe, S. M. & Wilson, J. R. Structural basis for the requirement of additional factors for MLL1 SET domain activity and recognition of epigenetic marks. *Mol. Cell* **33**, 181–191 (2009).
22. Verreault, A., Kaufman, P. D., Kobayashi, R. & Stillman, B. Nucleosome assembly by a complex of CAF-1 and acetylated histones H3/H4. *Cell* **87**, 95–104 (1996).
23. Murzina, N. V. *et al.* Structural basis for the recognition of histone H4 by the histone-chaperone RbAp46. *Structure* **16**, 1077–1085 (2008).
24. Schwartz, Y. B. & Pirrotta, V. Polycomb complexes and epigenetic states. *Curr. Opin. Cell Biol.* **20**, 266–273 (2008).
25. Ringrose, L. & Paro, R. Polycomb/Trithorax response elements and epigenetic memory of cell identity. *Development* **134**, 223–232 (2007).
26. Papp, B. & Muller, J. Histone trimethylation and the maintenance of transcriptional ON and OFF states by trxG and PcG proteins. *Genes Dev.* **20**, 2041–2054 (2006).
27. Schwartz, Y. B. *et al.* Genome-wide analysis of Polycomb targets in *Drosophila melanogaster*. *Nature Genet.* **38**, 700–705 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. McCabe for technical assistance, K. Basler for fly stocks and the pUASTattB plasmid, T. Jenuwein for antibodies and J. Muller for *Drosophila* PRC2 baculovirus. We thank J. Millar and A. Gould for suggestions and discussions, D. Allis for insightful discussions on histone specificity, P. Walker for technical assistance, and J. Brock for assistance with figures. This work was supported by the following grants: Fellowship from the Deutsche Akademie der Naturforscher Leopoldina (LPDS 2009-5) to P.V., NIH grants GM064844 and GM37120 and HHMI to D.R. Work in the S.J.G. laboratory is funded by the MRC. Work in the V.P. laboratory was supported by the Division of Life Sciences of Rutgers University.

Author Contributions R.M., N.J., K.O., M.L.S., J.S., W.J.D., P.V., S.R.M. and V.D.M. performed experiments. All authors contributed to data analysis, experimental design and manuscript writing.

Author Information The structural data have been deposited with the Protein Data Bank under accession numbers 3IJC (EED/ND5B), 3IIW (EED/H3K27), 3IIY (EED/H1K26), 3IJO (EED/H3K9) and 3IIJ (EED/H4K20). Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to V.P. (pirrotta@biology.rutgers.edu), D.R. (reinbd01@med.nyu.edu) or S.J.G. (sgambli@nimr.mrc.ac.uk).

METHODS

Protein expression and purification. Residues 78–441 of EED (Δ EED) were cloned into pGEX-4T vector (Amersham Biosciences) and expressed in *E. coli*. Proteins were prepared as N-terminal GST-fusion proteins and cleaved from GST with human α -thrombin (Haematologic Technologies, Inc.). Proteins were purified further using size-exclusion chromatography (Superdex 200, GE Healthcare) in buffer containing 50 mM Tris-HCl, pH 8.7, 150 mM NaCl and 3 mM TCEP. Site-directed mutants of Δ EED were generated with the ExSite protocol (Stratagene) and purified in a similar manner. Crystallographic and binding studies were carried out using a construct containing the mutation Met370Thr, however the binding properties of this construct are identical to those of the 'wild-type' construct. Peptides were synthesized and purified by reversed phase HPLC at the University of Bristol Peptide Synthesis Facility. Peptide masses were verified by mass spectrometry.

Crystallography. For crystallization trials, protein solutions were prepared either as Δ EED alone at 2.5 mg ml⁻¹ or as a complex solution at 1.5 mg ml⁻¹ with peptide at a sevenfold higher molar ratio. All protein solutions contained TCEP at 15 mM concentration. Crystals were grown at 18 °C using the vapour diffusion technique in hanging drops. Drops were prepared by mixing equal volumes of Δ EED protein alone with reservoir solution containing 4.0–4.1 M formate and 0.6–0.7 M NDSB-195, or by mixing equal volumes of Δ EED protein complex with 3.7–3.9 M formate solution. Crystals were transferred into mother liquor with 5–10% glycerol before flash cooling in liquid nitrogen. Diffraction data for the Δ EED-only native and selenomethionine crystals were collected at the Daresbury synchrotron on beamline 10.1 at the peak wavelength for selenium. Diffraction data for the H1K26me3, H3K9me3 and H4K20me3 protein complex crystals were collected using an in-house MicroMax 007HF rotating anode coupled to a RaxisIV⁺⁺ detector. Data for H3K27me3 was collected at Diamond Light Source on beamline I04 at a wavelength of 0.97 Å. Data were integrated using Denzo and scaled with Scalepack²⁸. Phases for the selenomethionine-substituted Δ EED structure were generated and extended using the SAD method and SOLVE²⁹ and RESOLVE³⁰ programs. Phases from RESOLVE were used to autobuild a model with ARP/wARP³¹ in warpNtrace mode. The protein complex crystal structures were solved by molecular replacement using AMoRe³² and the selenomethionine-substituted Δ EED structure as the search model. Standard refinement was carried out with refmac5 (ref. 33) and CNS³⁴ together with manual model building with O³⁵ and Coot³⁶. Figures were created with Pymol (DeLano Scientific; <http://pymol.sourceforge.net/>).

Binding studies. Histone-peptide-binding experiments were performed by competition fluorescence spectroscopy and ITC. All fluorescence emission spectra were measured using a dansyl-labelled peptide (sequence: KKKARK(Me3)SAGAAK-dansyl) at 20 °C in 50 mM Tris-HCl, pH 8.7, 150 mM NaCl and 3 mM TCEP. Measurements were recorded using a SPEX FluoroMax fluorimeter (excitation wavelength 330 nm, emission wavelength 537 nm). Binding of dansyl peptide to EED was monitored by titrating excess EED into 5 μ M peptide. Dissociation constants for the unlabelled histone peptides were determined using a competition assay by adding excess unlabelled peptide to a complex of 35 μ M EED with 35 μ M dansyl peptide and monitoring the subsequent reduction in fluorescence. ITC measurements were carried out by injecting peptide at 400–1,000 μ M into the ITC cell containing Δ 77EED at 40–100 μ M. Experiments were performed at 20 °C in 50 mM Tris-HCl, pH 8.7, 150 mM NaCl and 3 mM β -mercaptoethanol (BME).

Methyl lysine analogue production. Pseudo-lysine (ϕ K)-containing histones were generated by a modification of known literature methods¹². In brief, proteins to be modified (5–10 mg) were weighed into 1.5-ml siliconized Eppendorf microcentrifuge tubes and 950 μ l alkylation buffer (4 M guanidinium chloride, 1 M HEPES, 10 mM D/L-methionine at pH 7.8) the solution is passed through a 0.22- μ m filter and purged with argon before use) was added. Proteins that do not readily dissolve were sonicated for 10–15 min in a Branson 1510 ultrasonic cleaning bath at ambient temperature to affect dissolution. The resultant clear colourless solutions were treated with 20 μ l of a 1 M dithiothreitol (DTT) solution in alkylation buffer prepared just before use, and agitated at 37 °C for 1 h. At the end of this period the fully reduced proteins were treated as indicated below.

(1) Pseudo-lysine (ϕ K-NH₂): 100 μ l of a 1 M 2-chloroethylamine monohydrochloride solution in alkylation buffer (prepared just before use) was added to the reduced histone. The mixture was agitated in the dark at 45 °C for 2.5 h. At the end of this period the mixture was treated with a second portion of DTT (10 μ l of the above 1 M solution) and heated with agitation at 45 °C for a further 2.5 h. The reaction was then quenched with BME (50 μ l) and cooled to room temperature before purification as outlined below.

(2) Pseudo-monomethyl-lysine (ϕ K-Me1): 100 μ l of a 1 M N-methylaminoethyl chloride hydrochloride solution in alkylation buffer (prepared just before use) was

added to the reduced histone. The mixture was agitated in the dark at 45 °C for 2.5 h. At the end of this period the mixture was treated with a second portion of DTT (10 μ l of the above 1 M solution) and heated with agitation at 45 °C for a further 2.5 h. The reaction was then quenched with BME (50 μ l) and cooled to room temperature before purification as outlined below.

(3) Pseudo-dimethyl-lysine (ϕ K-Me2): 50 μ l of a 1 M 2-(dimethylamine)ethyl chloride hydrochloride solution in alkylation buffer (prepared just before use) was added to the reduced histone. The mixture was agitated in the dark at 25 °C for 2 h. At the end of this period the mixture was treated with more DTT (10 μ l of the above 1 M solution) and agitated at 25 °C for 30 min before addition of 50 μ l of the 1 M 2-(dimethylamine)ethyl chloride hydrochloride solution. The reaction was allowed to proceed at ambient temperature for a further 2 h, quenched with BME (50 μ l), and cooled to room temperature before purification as outlined below.

(4) Pseudo-trimethyl-lysine (ϕ K-Me3): 100 μ l of a 1 M (2-bromoethyl) trimethyl-ammonium bromide solution in alkylation buffer (prepared just before use) was added to the reduced histone. The mixture was agitated in the dark at 50 °C for 2.5 h. At the end of this period the mixture was treated with a second portion of DTT (10 μ l of the above 1 M solution) and heated with agitation at 50 °C for an extra 2.5 h. The reaction was then quenched with BME (50 μ l) and cooled to room temperature before purification as outlined below.

Purification scheme: A PD-10 column was pre-equilibrated with 0.1% BME in 18 Ω water. This was loaded with the reaction mixture, the reaction tube was rinsed with 1 ml alkylation buffer and this was also added to the top of the column. The proteins were then eluted according to the manufacturer's protocol for centrifugal isolation. The eluent was frozen and lyophilized before providing the modified histones as crispy foams. A portion of each (~0.1 mg) was analysed by reverse-phase-HPLC and matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectrometry to ensure product identity and homogeneity.

Chromatin and interaction experiment. Histone H3 variants with the respective point mutations (Lys to Cys at the position to be modified, and Cys to Ala at position 110) were expressed in *E. coli*, purified from inclusion bodies, and solubilized in 7 M guanidine hydrochloride, 20 mM Tris, pH 8, 10 mM DTT. After dialysis to replace guanidine hydrochloride with 7 M urea, histones were further purified by sequential anion and cation chromatography. Histone-containing fractions were pooled, dialysed against 5 mM BME, and lyophilized. Histones were reconstituted into octamer as previously described³⁷ and chromatin was formed by salt dialysis. To prevent unspecific binding to free histone in the pull-down experiment, chromatin was further purified on an agarose2 column. For interaction, 2 μ g chromatin was incubated with 2 μ g protein or complex of interest in buffer A (50 mM Tris, pH 8.0, 50 mM NaCl, 1 mM EDTA, 0.01% NP40) for 2 h at 4 °C in the presence of Ni-NTA beads (EED) or M2-beads (PRC2). Beads were extensively washed, eluted with 1 \times SDS-PAGE loading buffer and analysed by western blot.

HKMT assay. HKMT assays were performed as previously described³⁸. For autoradiography exposure, the conditions were as follows: 1.5 μ g chromatin, 100 ng reconstituted PRC2 complex, 5–40 μ M peptide and 0.3 μ M ³H-SAM. For scintillation counting, the assay was performed as follows: 1.5 μ g chromatin, 50 ng reconstituted PRC2 complex, 100 μ M peptide, 24.8 μ M SAM (³H-SAM/SAM ratio 1/30) for 15 min, unless otherwise stated in the figure legend.

SF9 culture, infection and complex purification. As previously described³⁸.

Antibodies. H3K27me1 (Millipore), H3K27me2 (Abcam, ab24684 and ab6002), H3K27me2 and H3K27me3 (gift from T. Jenuwein), total H3 (Abcam, ab1791), Flag (Sigma), Myc 9E10 (Chemicon). Previously described antibodies were used for EED³⁸ and E(Z)³¹.

Fly strains and mutants. The *Df(1)y¹w^{67c23}* strain (*yw*) was used as wild-type control and for P-mediated germ-line transformation. For transgene insertion at specific genomic sites, we used fly strains in which the ϕ C31 integrase gene is inserted on the X chromosome and attP landing sites are located in 68E or 86Fb, gifts from K. Basler³⁹. Mutant strains for *esc^c* and *esc^c esc^d*^{d01514} were used as described previously¹⁵, and detailed crossing schemes to test transgene function are given in Supplementary Figs 9 and 10. ChIP with larval tissues was done using flies homozygous for *esc^c*>Myc-ESC in homozygous *esc^c esc^d*^{d01514} background or flies homozygous for *esc^c*>Myc-ESC(Phe77Ala) in homozygous *esc^c esc^d*^{d01514} (selected using a *CyO*, GFP balancer).

Transposon construction. ESC mutant transgenes were produced using the *esc^c*>Myc-ESC construct¹⁵ as starting material. The *esc^c*>Myc-ESC(Phe77Ala) and *esc^c*>Myc-ESC(Tyr338Ala) constructs for conventional P-mediated germ-line transformation were assembled in the pCaS-*escp* construct¹⁵. To generate transgenic lines for *esc^c*>Myc-ESC or *esc^c*>Myc-ESC(Phe345Ala) at the same chromosomal locations, we used the ϕ C31 recombinase-mediated cassette exchange technique (RMCE)⁴⁰. For recombinase-mediated cassette exchange, pCaSpeRattB plasmid was first generated by excising with BamHI the UAS and *hsp70* minimal promoter from pUASTattB (a gift from K. Basler) and

replacing them with a PCR-amplified multicloning site with BglII cohesive ends. A fragment containing the *esc* promoter was excised from the *esc>Myc-ESC* construct by NotI and KpnI and inserted in the NotI–KpnI site of pCaSpeRattB to generate pCaB-*escp*. The *esc>Myc-ESC* and *esc>Myc-ESC(Phe345Ala)* constructs for RMCE were constructed by inserting the wild-type *esc* cDNA or *esc* cDNA with the relevant mutation in the KpnI site of the pCaB-*escp*.

Western blotting and immunoprecipitation. *Drosophila* larval extracts were prepared as previously described¹⁵. In brief, approximately 30 third instar larvae, frozen in liquid nitrogen, were pulverized in protein lysis buffer I (50 mM Tris, pH 6.8, 100 mM DTT, 2% SDS, 5 mM EDTA, 1 mM phenylmethyl sulphonyl fluoride (PMSF) and 10% glycerol) using a micropestle. After heating at 95 °C for 5 min, and centrifuging for 10 min, supernatants were used for western blot analysis. Ovary extracts were made by homogenizing 60 sets of ovaries in extraction buffer (30 mM HEPES-potassium hydroxide, pH 7.6, 150 mM potassium acetate, 2 mM magnesium acetate, 5 mM DTT, 0.1% NP40 and Protein Inhibitor cocktail (Roche)) using a micropestle, and the extracts were cleared by centrifugation. Immunoprecipitation was performed using anti-Myc and protein G sepharose beads (GE Healthcare). The beads were washed five times for 5 min in extraction buffer, boiled in sample buffer and analysed by western blotting. Rabbit anti-H3 and rabbit anti-H3K27me2 or -H3K27me3 were used with anti-rabbit IgG-alkaline-phosphatase as a secondary antibody. The rabbit anti-E(Z) was used with goat anti-rabbit IgG-horseradish peroxidase (HRP) and mouse anti-Myc was used with goat anti-mouse IgG–HRP light-chain-specific.

ChIP with *Drosophila* larvae. ChIP was performed essentially as previously described with slight modifications³⁰. Approximately 300 mg of larvae were taken for two independent experiments. The frozen larvae were first pulverized using a mortar and pestle in liquid N₂, then homogenized with 10 strokes of a Dounce homogenizer in cross-linking solution (1.8% formaldehyde, 50 mM HEPES, pH 8.0, 1 mM EDTA, 0.5 mM EGTA, 100 mM NaCl). The homogenates were incubated by rotating at room temperature for 15 min. The fixation was stopped by washing for 5 min in 0.01% Triton X-100, 0.125 M glycine in PBS three times with mixing. After centrifugation at 1,500g for 3 min at 4 °C, the pellets were washed for 10 min in 1 ml of wash buffer A (10 mM HEPES, pH 7.6, 10 mM EDTA, 0.5 mM EGTA and 0.25% Triton X-100) and subsequently for 10 min in 1 ml wash buffer B (10 mM HEPES, pH 7.6, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA and 0.01% Triton X-100) by mixing gently. The washed pellets were resuspended in sonication buffer (10 mM HEPES, pH 7.6, 1 mM EDTA and 0.5 mM EGTA). Sonication was performed with a Bioruptor as described previously³¹. After sonication, samples were supplemented with *N*-lauroylsarcosine (0.5% final) and incubated for 10 min at 4 °C with gentle mixing. Soluble chromatin was fractionated by centrifugation at top speed for 10 min and transferred to new eppendorf tubes. The chromatin was aliquoted, quick-frozen in liquid N₂ and stored at –80 °C. All steps for immunoprecipitation were performed at 4 °C. An aliquot of sonicated chromatin was first pre-cleared by mixing with Protein G Sepharose beads (GE Healthcare) or Protein A Sepharose beads (Sigma) for 1 h. After centrifugation, pre-cleared chromatin was incubated with anti-Myc9E10 (Chemicon), anti-E(Z) or anti-H3K27me3 (Abcam, ab6002) overnight. Protein G or protein A sepharose beads were added to allow binding to the antibody for 2 h and then washed five times with RIPA

buffer, once with LiCl buffer (10 mM Tris-HCl, pH 8.0, 250 mM LiCl, 0.5% NP-40, 0.5% sodium deoxycholate and 1 mM EDTA) and twice with TE buffer. The beads were resuspended in 100 µl TE buffer, and treated with 0.1 mg ml^{–1} RNase A at 37 °C for 30 min. After supplementing with SDS (0.5% final), the beads were treated with 0.5 mg ml^{–1} Proteinase K at 37 °C overnight and subsequently at 65 °C for 6 h. The immunoprecipitated DNA was recovered by phenol–chloroform extraction and ethanol precipitation, and then dissolved in H₂O. Control mock immunoprecipitations were done in the same way except that no antibodies were added to the reaction mixture. Real-time PCR quantification of immunoprecipitated DNA was performed as previously described³¹. The input DNA extracted from the same sonicated chromatin aliquots as above was used to plot a standard curve. Primers were as follow: for *bx*d PRE (FM4 and FM6), FM4.1, 5'-AGCAATTTGTCACCGCAAGG-3', FM4.2, 5'-GGATTTTGAGTGC GTTCTTCC-3', FM6.1, 5'-CCAAACGGAAAAGCGAGTGG-3', and FM6.2, 5'-GCACTAAACCCCATAAAAGTC-3'; for PBX enhancer, PBX-enh-5', 5'-GA AAACACACAAGTGCAG-3' and PBX-enh-3', 5'-GGAGATCCTAAACAT GC-3'; for *Ubx* promoter, U-up1.1, 5'-ATTGCGGAGATACCAATGCC-3' and U-up1.2, 5'-ATTGCGGAGATACCAATGCC-3'; for *white* locus, W2.1, 5'-ATGCCACGACATCTGACC-3' and w2.3, 5'-AATGCCAGACGCTTCCT TC-3'. The quantity obtained by real-time PCR was corrected to obtain the percentage of input.

28. Otwinowski, Z. & Minor, W. Processing of x-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **267**, 307–326 (1997).
29. Terwilliger, T. C. & Berendzen, J. Automated MAD and MIR structure solution. *Acta Crystallogr. D* **55**, 849–861 (1999).
30. Terwilliger, T. C. Maximum-likelihood density modification. *Acta Crystallogr. D* **56**, 965–972 (2000).
31. Perrakis, A., Sixma, T. K., Wilson, K. S. & Lamzin, V. S. wARP: improvement and extension of crystallographic phases by weighted averaging of multiple-refined dummy atomic models. *Acta Crystallogr. D* **53**, 448–455 (1997).
32. Navaza, J. AMoRe: an automated package for molecular replacement. *Acta Crystallogr. A* **50**, 157–163 (1994).
33. Collaborative Computational Project. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
34. Brünger, A. T. et al. Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
35. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
36. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
37. Luger, K., Rechsteiner, T. J., Flaus, A. J., Wayne, M. M. & Richmond, T. J. Characterization of nucleosome core particles containing histone proteins made in bacteria. *J. Mol. Biol.* **272**, 301–311 (1997).
38. Margueron, R. et al. Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms. *Mol. Cell* **32**, 503–518 (2008).
39. Bischof, J., Maeda, R. K., Hediger, M., Karch, F. & Basler, K. An optimized transgenesis system for *Drosophila* using germ-line-specific phiC31 integrases. *Proc. Natl Acad. Sci. USA* **104**, 3312–3317 (2007).
40. Bateman, J. R., Lee, A. M. & Wu, C. T. Site-specific transformation of *Drosophila* via phiC31 integrase-mediated cassette exchange. *Genetics* **173**, 769–777 (2006).

LETTERS

Quantum signatures of chaos in a kicked top

S. Chaudhury¹, A. Smith¹, B. E. Anderson¹, S. Ghose² & P. S. Jessen¹

Chaotic behaviour is ubiquitous and plays an important part in most fields of science. In classical physics, chaos is characterized by hypersensitivity of the time evolution of a system to initial conditions. Quantum mechanics does not permit a similar definition owing in part to the uncertainty principle, and in part to the Schrödinger equation, which preserves the overlap between quantum states. This fundamental disconnect poses a challenge to quantum–classical correspondence¹, and has motivated a long-standing search for quantum signatures of classical chaos^{2,3}. Here we present the experimental realization of a common paradigm for quantum chaos—the quantum kicked top^{2,4}—and the observation directly in quantum phase space of dynamics that have a chaotic classical counterpart. Our system is based on the combined electronic and nuclear spin of a single atom and is therefore deep in the quantum regime; nevertheless, we find good correspondence between the quantum dynamics and classical phase space structures. Because chaos is inherently a dynamical phenomenon, special significance attaches to dynamical signatures such as sensitivity to perturbation^{1,5} or the generation of entropy⁶ and entanglement^{7,8}, for which only indirect evidence has been available^{9–11}. We observe clear differences in the sensitivity to perturbation in chaotic versus regular, non-chaotic regimes, and present experimental evidence for dynamical entanglement as a signature of chaos.

In classical mechanics, the state of a physical system is specified by a set of dynamical variables—for example, the position and momentum of a point particle—whose values define a point in phase space. Regular motion is associated with periodic orbits in phase space, whereas chaos is characterized by complex, aperiodic trajectories that diverge exponentially as a function of initial separation. This description of states and time evolution is fundamentally incompatible with quantum mechanics, where conjugate observables such as position and momentum cannot take on well-defined values at the same time. However, it is still possible to represent a quantum state in phase space, in the form of a delocalized quasi-probability distribution whose evolution is governed by the Schrödinger equation¹². This suggests an experiment in which one prepares an initial minimum uncertainty state centred on a point in phase space, subjects it to a desired evolution, measures the quantum state at successive points in time, and observes the degree to which the dynamically evolving quantum phase space distribution reflects the classical phase space structures. Experiments of this type can be simulated with classical waves³, but are very challenging for true quantum systems because of the overhead involved in state preparation, control and reconstruction. Quantum versions that accomplish several of the steps have been performed with cold atoms in laser standing waves^{13,14}, and in this Letter we complete the entire programme by including full quantum state reconstruction and visualizing the dynamics via complete phase space distributions. Placing the emphasis on dynamics complements the much larger body of experimental work on energy level statistics in a broad range of physical systems^{15–17}.

The experimental tools required to study quantum chaos directly in phase space have recently become available for the physical system

consisting of the spin angular momentum of a single ¹³³Cs atom in the $F = 3$ hyperfine ground state^{18–20}. To take advantage of this, we have implemented a very popular model system known as the ‘kicked top’, consisting of a spin F whose dynamics is governed by a periodic Hamiltonian:

$$H = \hbar p F_y \sum_{n=0}^{\infty} f(t - n\tau) + \hbar \frac{\kappa}{2F\tau} F_x^2 \quad (1)$$

In the simplest case, the kick f is a δ -function, and each period of the classical evolution breaks down into a rotation about the y axis by a fixed angle p , followed by a twist (a rotation about the x axis by an angle proportional to F_x). The parameter κ determines the degree to which the dynamics are regular or chaotic. In our experiment, the rotation is performed by applying a short magnetic field pulse, whereas the twist is induced by the a.c. Stark shift (light shift) from a laser field tuned near the D1 resonance at 895 nm (ref. 18). Because the magnitude of the spin is conserved, phase space is a spherical surface on which each point represents a particular orientation of the spin, and the classical evolution can be visualized by a stroboscopic plot showing the state at times $t = n\tau$. Figure 1 shows such a phase space plot for our kicked top, with parameters $p = 0.99$ and $\kappa = 2.0$. We see immediately that the phase space is mixed, with one large island of regular motion in the $F_y < 0$ hemisphere, two smaller islands in the $F_y > 0$ hemisphere, and a sea of chaos almost everywhere else.

To visualize a quantum state of the kicked top in phase space, one can expand it in a basis of spin-coherent states $|\theta, \phi\rangle$, which are minimum uncertainty states with maximum projection in the directions

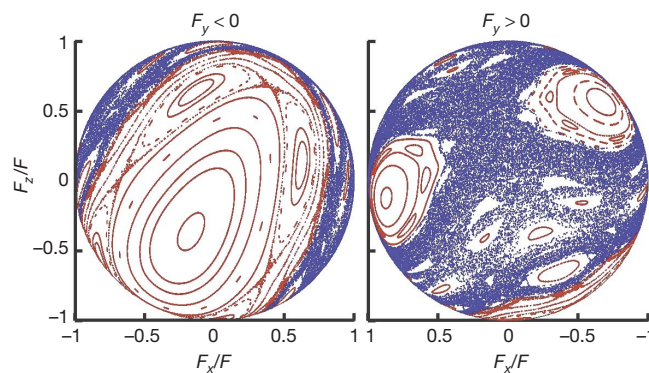


Figure 1 | Stroboscopic phase space plot for a classical kicked top.

Trajectories are obtained by integrating the classical equations of motion and plotting the position of the spin after each kick. Depending on the starting point, states follow regular orbits (red), or move along chaotic trajectories (blue). Motion across the boundaries between regular and chaotic regions is classically forbidden. For this plot $p = 0.99$ and $\kappa = 2.0$, resulting in a mixed phase space that contains both regular islands of various sizes and a sea of chaos. The $F_y < 0$ and $F_y > 0$ hemispheres are shown separately, respectively left and right.

¹College of Optical Sciences, University of Arizona, Tucson, Arizona 85721, USA. ²Department of Physics and Computer Science, Wilfrid Laurier University, Waterloo, Ontario N2L 3C5, Canada.

given by the polar and azimuthal angles (θ, ϕ) and thus the closest quantum approximation to a classical spin. This produces the Husimi quasi-probability distribution, $Q(\theta, \phi) = (2F + 1) \langle \theta, \phi | \rho | \theta, \phi \rangle / 4\pi$, where ρ is the density operator for the quantum state (pure or statistical mixture)²¹. $Q(\theta, \phi)$ is a normalized, everywhere positive function that comes as close as possible to a classical probability distribution in phase space.

We use as a starting point for our kicked-top experiments an ensemble of laser-cooled Cs atoms prepared by optical pumping in a desired spin-coherent state $\rho_0 \approx |\theta, \phi\rangle\langle\theta, \phi|$. In a given run of the experiment, each member of the ensemble is subjected to n periods of the kicked-top Hamiltonian, and the entire density operator for the final state is experimentally reconstructed¹⁹. The process is repeated for $0 \leq n \leq 40$, in order to build up a stroboscopic record $\{\rho_n\}$ for the evolving quantum state. Finally, we carry out the entire procedure for a series of initial states. To obtain a visual quantum–classical comparison, we convert each data set $\{\rho_n\}$ into Husimi distributions to obtain a ‘stop-motion movie’ of the evolution of the state. Figure 2A shows selected frames from a movie obtained for an initial spin-coherent state centred on the stable island near $F_x/F = 1$ in the $F_y > 0$ hemisphere. Successive frames clearly show the phenomenon of dynamical tunnelling, wherein the quantum system oscillates between two regions of phase space even though motion through the chaotic sea is classically forbidden^{13,14}. The observed tunnelling period is in good agreement with a prediction based on decomposition of the initial state into Floquet eigenstates (Supplementary

Information). It is also clear that the tunnelling oscillation dephases after roughly one period. This is a sign of imperfections in our experiments, mainly a 5% variation in κ due to laser intensity variation across the ensemble, decoherence induced by spontaneous light scattering (~ 1 photon per 15 kicks), and background magnetic fields. Our data are in good quantitative agreement with a full master equation calculation that includes these imperfections.

An additional, useful visualization of data of the type displayed in Fig. 2 can be achieved by averaging the Husimi distribution over many cycles. The result is a single plot showing the parts of phase space accessible from a given initial state. Figure 2C shows 40-period averages for three initial states, which together illustrate the remarkable degree to which our quantum kicked top reflects the boundaries between regular and chaotic regions in classical phase space. Although this is to be expected for systems in the mesoscopic regime, it is somewhat surprising that our deeply quantum mechanical spin should do so.

In recent years, much attention has been directed towards dynamical signatures of chaos in quantum systems. One candidate is the sensitivity to perturbation, which can be quantified by the decay in overlap between quantum states evolving according to two slightly different Hamiltonians, and which can potentially reflect the classical Lyapunov exponent²². The spin in our experiment is too small for the overlap to undergo exponential decay, but different sensitivities to perturbation should still be reflected in the decay of the purity of the spin density operator, as each spin evolves with a slightly different value of κ , and is coupled to the environment through light scattering.

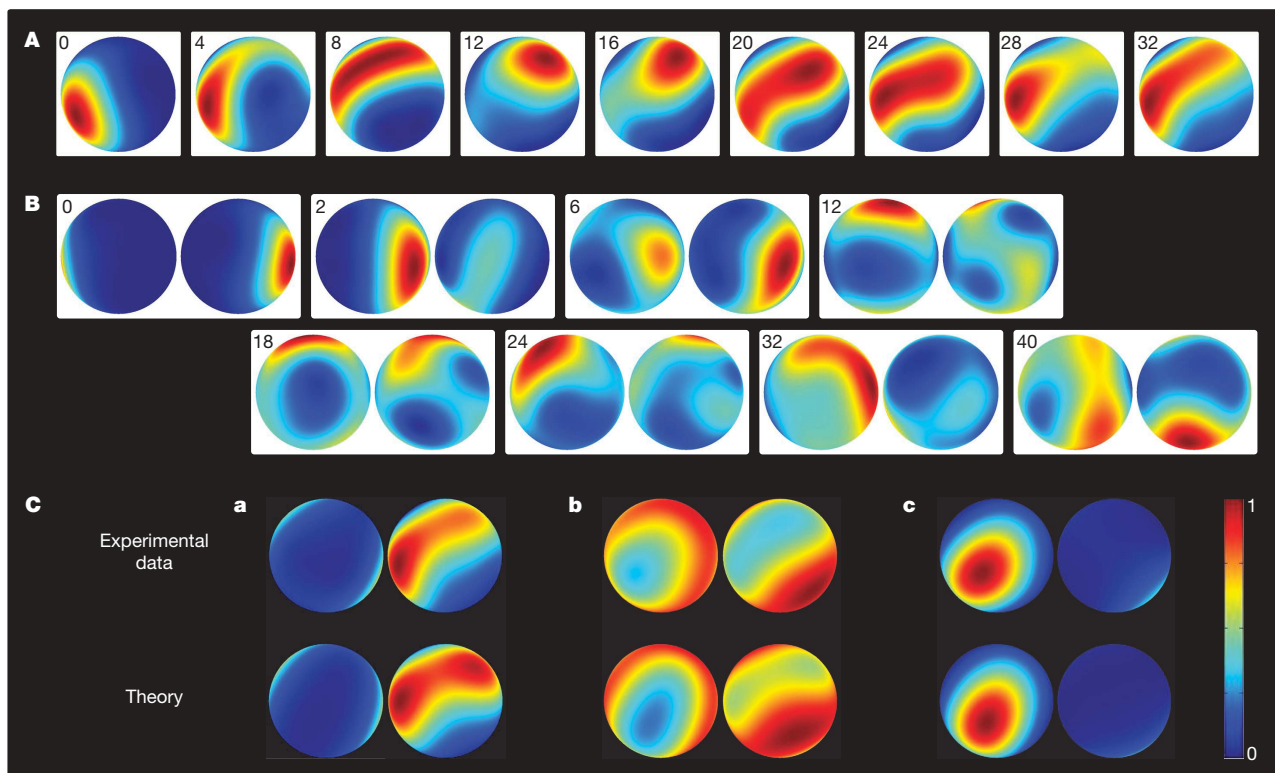


Figure 2 | Quantum phase space (Husimi) distributions for a quantum kicked top. Stroboscopic illustration of dynamical evolution, showing selected experimental snapshots from the first 40 periods of the kicked-top Hamiltonian. The period number is indicated in each frame. **A**, The initial state is a spin-coherent state centred at $F_x/F = 0.70$, $F_y/F = 0.70$, $F_z/F = -0.16$, where it is mostly contained within the boundaries of the lower of the pair of islands in the $F_y > 0$ hemisphere of Fig. 1. The state undergoes roughly 1.5 periods of dynamical tunnelling before coherence is lost. The state is almost entirely confined to the $F_y > 0$ hemisphere, which is the only one shown. **B**, The initial state is a spin-coherent state centred at $F_x/F = -0.94$, $F_y/F = 0.31$, $F_z/F = -0.16$, where it is mostly contained within the chaotic sea. The state spreads into the chaotic regions but generally avoids the regular islands. Both hemispheres are shown.

C, 40-period averages of evolving Husimi distributions. **a–c**, The initial states are spin-coherent states centred at **a**, $F_x/F = 0.70$, $F_y/F = 0.70$, $F_z/F = -0.16$ (island in the $F_y > 0$ hemisphere, same as **A**), at **b**, $F_x/F = -0.94$, $F_y/F = 0.31$, $F_z/F = -0.16$ (in the chaotic sea, same as **B**), and at **c**, $F_x/F = 0$, $F_y/F = -0.99$, $F_z/F = -0.16$ (large island in the $F_y < 0$ hemisphere). The upper data set is the observation from experiments, while the lower data set is the prediction of a full theoretical model taking into account decoherence and κ variation. In combination, the quantum phase space distributions reflect the classical phase space structures of Fig. 1 with remarkable fidelity. To enhance contrast, the Husimi distribution in each image has been rescaled to fit the interval $[0, 1]$. Each quantum state is experimentally reconstructed with a fidelity of $\sim 90\%$.

Figure 3 shows the experimentally measured state purity, $\text{Tr}[\rho^2]$, as a function of period number, for two different initial states. As predicted, the purity decays at very different rates in regular and chaotic regions.

For systems with multiple degrees of freedom, it has been argued that classical chaos is linked to the dynamical generation of entanglement in the quantum system^{23,24}. Our atomic spin is the sum of electron and nuclear spins, $\mathbf{F} = \mathbf{S} + \mathbf{I}$. It is therefore natural to test if the entanglement generated between the two is a reliable signature of chaos. Here we use the linear entropy $S_{\text{LE}} = 1 - \text{Tr}[\rho_e^2]$ as our entanglement measure, where ρ_e is the reduced density operator for the electron spin. This is reasonable as long as the overall state is nearly pure. In our experiment, S_{LE} reaches steady state after just a few kicks (Supplementary Information), and the 40-period average therefore serves as a convenient and robust measure of the entanglement generated by the dynamics. Figure 4 shows a significant dip in $\langle S_{\text{LE}} \rangle$ and correspondingly less entanglement generation for initial states localized in regular regions compared to those in the chaotic sea. This is (to our knowledge) the first experimental evidence that the purely quantum property of entanglement is a good signature of classical chaos. Note that whereas the signature is very clear for initial states in the large regular island in the $F_y < 0$ hemisphere, it is less apparent for states located on the small island in the $F_y > 0$ hemisphere. This loss in contrast occurs because the latter become entangled by dynamical tunnelling, and is therefore linked to the deeply quantum nature of our small spin. Contrast is further reduced by the sensitivity of tunnelling to κ variations and decoherence, which is apparent from the difference between our perturbation-free and full models (Supplementary Information). Tunnelling will be suppressed for much larger spins, and it is reasonable to assume that the distinction between regular and chaotic regions will be more universal in that regime.

Our laboratory realization of the kicked top with atomic spins points the way to further studies of quantum chaos in the time domain. We are currently working to extend our control and measurement tools to the entire hyperfine ground manifold of the Cs atom²⁵, which will provide access to the full state space for the coupled electron–nuclear spins. Besides increasing the size of state space by more than a factor of two, this will offer a more powerful platform for the study of entanglement as a quantum signature of chaos²⁶. To reach the semiclassical limit of very large spins, one can in principle implement a kicked-top Hamiltonian for the collective spin of an atomic ensemble, with the twisting interaction induced either by ultracold collisions²⁷, or by coupling the spins to a shared mode of a quantized electromagnetic field²⁸.

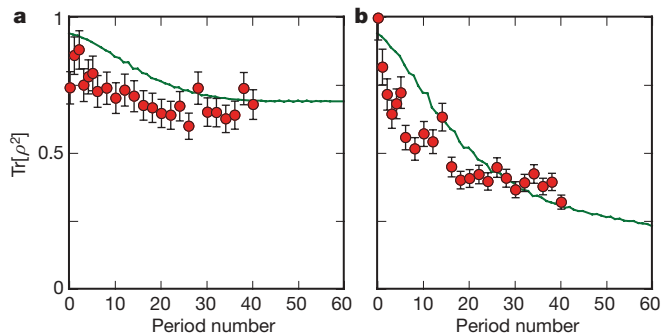


Figure 3 | Sensitivity to perturbation as a quantum signature of chaos. The purity of the spin density operator, $\text{Tr}[\rho^2]$, shown as a function of the period number. **a**, Initial state localized on the large island in the $F_y < 0$ hemisphere (same data set as in Fig. 2C, c). **b**, Initial state localized in the sea of chaos (same data set as in Fig. 2C, b). Red circles are experimental data and the green lines are the predictions of a full model. As expected, perturbations, in the form of decoherence and κ variation across the ensemble, reduce the purity much faster for a state in the chaotic sea. Experimental error bars, ± 1 s.d.

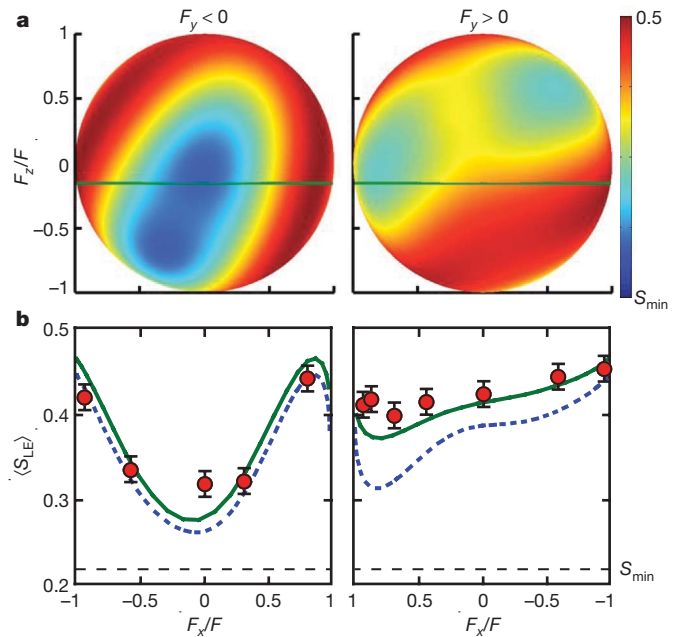


Figure 4 | Entanglement as a quantum signature of chaos. Entanglement between the electron and nuclear spins is quantified by the linear entropy, $S_{\text{LE}} = 1 - \text{Tr}[\rho_e^2]$, of the electron reduced density operator. It is averaged over 40 periods of the kicked-top Hamiltonian, and shown as a function of the centre of the initial spin coherent state $|\theta, \phi\rangle$. **a**, Theoretical prediction for Schrödinger evolution, corresponding to an ideal situation without perturbations (no decoherence or κ variation). Colours indicate the value of $\langle S_{\text{LE}} \rangle$. **b**, Experimental measurements performed for states lying along the green cross-section in **a**. Also shown are the predictions of a full model (solid green line) and the perturbation free model (dashed blue line) used in **a**. The black dashed line is the linear entropy of a minimally entangled pure state in the $F = 3$ manifold. A marked contrast in dynamically generated entanglement can be seen between regular and chaotic regions. Experimental error bars, ± 1 s.d.

Ultimately, this could allow experiments to address some of the most important outstanding questions related to quantum–classical correspondence, such as how to recover classical (chaotic) dynamics in open quantum systems subject to decoherence²² or weak measurement²⁹.

METHODS SUMMARY

We prepare a spin ensemble by capturing and laser cooling $\sim 10^7$ Cs atoms in a magneto-optical trap and optical molasses. The atoms are released into free fall, and optically pumped into an $F = 3$ spin-coherent state with respect to a fixed axis. A set of precision coils driven by arbitrary waveform generators apply time-dependent magnetic fields in a bandwidth of ~ 200 kHz, and generate fast and accurate rotations through the magnetic interaction $g_F \mu_B \mathbf{B} \cdot \mathbf{F}$, where g_F is the Landé g factor for the spin \mathbf{F} and μ_B is the Bohr magneton. We use magnetic field pulses to prepare spin-coherent states along desired directions, and to perform the rotation in the kicked-top Hamiltonian. The continuous twist is induced by the a.c. Stark shift in a linearly polarized, monochromatic laser field, leading to an effective ground state Hamiltonian of the form $\hbar \xi F_x^2$ (ref. 18). In our experiment the magnetic kick duration is 17 μs , the peak Larmor frequency is 15 kHz, the twisting strength $\xi = 2\pi \times 533$ Hz, and the kicked-top period is $\tau = 100 \mu\text{s}$ for $\kappa = 2.0$. The finite duration of the magnetic kick causes overlap of the rotation and twisting parts of the evolution, but this does not significantly alter the character of the dynamics and is easily taken into account in the equations of motion. The rotation, $p = 0.99$, is chosen to maximize the size of the islands in the $F_y > 0$ hemisphere, and to allow a spin coherent state to be contained mostly within one of these.

Following n periods of the kicked-top Hamiltonian, the information needed to reconstruct the final spin density operator with a fidelity of $\sim 90\%$ is acquired during a single 2-ms phase of continuous weak optical measurement and dynamical control. Details regarding nonlinear spin dynamics, quantum control and quantum state reconstruction, and the theoretical modelling of spin dynamics, can be found in previous work published by our group^{18,19,30}.

Received 16 June; accepted 6 August 2009.

1. Peres, A. *Quantum Theory: Concepts and Methods* (Springer, 1995).
2. Haake, F. *Quantum Signatures of Chaos* (Springer, 2001).
3. Stöckmann, H.-J. *Quantum Chaos: An Introduction* (Cambridge Univ. Press, 1999).
4. Haake, F., Kus, M. & Scharf, R. Classical and quantum chaos for a kicked top. *Z. Phys. B* **65**, 381–395 (1987).
5. Peres, A. Stability of quantum motion in chaotic and regular systems. *Phys. Rev. A* **30**, 1610–1615 (1984).
6. Zurek, W. H. & Paz, J. P. Decoherence, chaos, and the Second Law. *Phys. Rev. Lett.* **72**, 2508–2511 (1994).
7. Furuya, K., Nemes, M. C. & Pellegrino, G. Q. Quantum dynamical manifestation of chaotic behavior in the process of entanglement. *Phys. Rev. Lett.* **80**, 5524–5527 (1998).
8. Jalabert, R. A. & Pastawski, H. M. Environment-independent decoherence rate in classically chaotic systems. *Phys. Rev. Lett.* **86**, 2490–2493 (2001).
9. Weinstein, Y. S., Lloyd, S., Emerson, J. & Cory, D. G. Experimental implementation of the quantum baker's map. *Phys. Rev. Lett.* **89**, 157902 (2002).
10. Ryan, C. A., Emerson, J., Poulin, D., Negrevergne, C. & Laflamme, R. Characterization of complex quantum dynamics with a scalable NMR information processor. *Phys. Rev. Lett.* **95**, 250502 (2005).
11. Andersen, M. F., Kaplan, A. & Davidson, N. Echo spectroscopy and quantum stability of trapped atoms. *Phys. Rev. Lett.* **90**, 023001 (2003).
12. Lee, H. W. Theory and application of the quantum phase-space distribution functions. *Phys. Rep.* **259**, 147–211 (1995).
13. Steck, D. A., Oskay, W. H. & Raizen, M. G. Observation of chaos-assisted tunneling between islands of stability. *Science* **293**, 274–278 (2001).
14. Hensinger, W. K. *et al.* Dynamical tunnelling of ultracold atoms. *Nature* **412**, 52–55 (2001).
15. Weidenmüller, H. A. & Mitchell, G. E. Random matrices and chaos in nuclear physics: nuclear structure. *Rev. Mod. Phys.* **81**, 539–589 (2009).
16. Blümel, R. & Reinhardt, W. P. *Chaos in Atomic Physics* (Cambridge Univ. Press, 1997).
17. Alhassid, Y. The statistical theory of quantum dots. *Rev. Mod. Phys.* **72**, 895–968 (2000).
18. Chaudhury, S. *et al.* Quantum control of the hyperfine spin of a Cs atom ensemble. *Phys. Rev. Lett.* **99**, 163002 (2007).
19. Smith, G. A., Silberfarb, A., Deutsch, I. H. & Jessen, P. S. Efficient quantum-state estimation by continuous weak measurement and dynamical control. *Phys. Rev. Lett.* **97**, 180403 (2006).
20. Ghose, S., Stock, R., Jessen, P., Lal, R. & Silberfarb, A. Chaos, entanglement, and decoherence in the quantum kicked top. *Phys. Rev. A* **78**, 042318 (2008).
21. Agarwal, G. S. Relation between atomic coherent-state representation, state multipoles, and generalized phase-space distributions. *Phys. Rev. A* **24**, 2889–2896 (1981).
22. Habib, S., Shizume, K. & Zurek, W. H. Decoherence, chaos, and the correspondence principle. *Phys. Rev. Lett.* **80**, 4361–4365 (1998).
23. Prosen, T. Chaos and complexity of quantum motion. *J. Phys. A* **40**, 7881–7918 (2007).
24. Jacquod, Ph & Petitjean, C. Decoherence, entanglement and irreversibility in quantum dynamical systems with few degrees of freedom. *Adv. Phys.* **58**, 67–196 (2009).
25. Merkel, S. T., Jessen, P. S. & Deutsch, I. H. Quantum control of the hyperfine-coupled electron and nuclear spins in alkali-metal atoms. *Phys. Rev. A* **78**, 023404 (2008).
26. Trail, C. M., Madhok, V. & Deutsch, I. H. Entanglement and the generation of random states in the quantum chaotic dynamics of kicked coupled tops. *Phys. Rev. E* **78**, 046211 (2008).
27. Micheli, A., Jaksch, D., Cirac, J. I. & Zoller, P. Many-particle entanglement in two-component Bose-Einstein condensates. *Phys. Rev. A* **67**, 013607 (2003).
28. Takeuchi, M. *et al.* Spin squeezing via one-axis twisting with coherent light. *Phys. Rev. Lett.* **94**, 023003 (2005).
29. Bhattacharya, T., Habib, S. & Jacobs, K. Continuous quantum measurement and the emergence of classical chaos. *Phys. Rev. Lett.* **85**, 4852–4855 (2000).
30. Silberfarb, A., Jessen, P. S. & Deutsch, I. H. Quantum state reconstruction via continuous measurement. *Phys. Rev. Lett.* **95**, 030402 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank I. H. Deutsch and P. Jacquod for discussions. This work was supported by the National Science Foundation (grant no. 0653631) and the Office of Naval Research (grant no. N00014-05-1-420). S.G. was supported by an NSERC Discovery grant.

Author Contributions All authors contributed extensively to this work.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to P.S.J. (poul.jessen@optics.arizona.edu).

LETTERS

Observation of unidirectional backscattering-immune topological electromagnetic states

Zheng Wang^{1*}, Yidong Chong^{1†*}, J. D. Joannopoulos¹ & Marin Soljačić¹

One of the most striking phenomena in condensed-matter physics is the quantum Hall effect, which arises in two-dimensional electron systems^{1–4} subject to a large magnetic field applied perpendicular to the plane in which the electrons reside. In such circumstances, current is carried by electrons along the edges of the system, in so-called chiral edge states (CESs). These are states that, as a consequence of nontrivial topological properties of the bulk electronic band structure, have a unique directionality and are robust against scattering from disorder. Recently, it was theoretically predicted^{5–7} that electromagnetic analogues of such electronic edge states could be observed in photonic crystals, which are materials having refractive-index variations with a periodicity comparable to the wavelength of the light passing through them. Here we report the experimental realization and observation of such electromagnetic CESs in a magneto-optical photonic crystal⁷ fabricated in the microwave regime. We demonstrate that, like their electronic counterparts^{8–13}, electromagnetic CESs can travel in only one direction and are very robust against scattering from disorder; we find that even large metallic scatterers placed in the path of the propagating edge modes do not induce reflections. These modes may enable the production of new classes of electromagnetic device and experiments that would be impossible using conventional reciprocal photonic states alone. Furthermore, our experimental demonstration and study of photonic CESs provides strong support for the generalization and application of topological band theories to classical and bosonic systems, and may lead to the realization and observation of topological phenomena in a generally much more controlled and customizable fashion than is typically possible with electronic systems.

The existence of photonic CESs was first predicted^{5,6} by an analogy between a photonic crystal^{14–16} with broken time-reversal symmetry and a system exhibiting the quantum Hall effect (QHE). In this analogy, the electromagnetic fields play the part of the electronic current, the variations of permittivity and permeability within the photonic crystal play the part of the periodic potential and the gradients of the gyrotropic components of the permeability tensor play the part of the external d.c. magnetic field that breaks the time-reversal symmetry^{5–7}. The defining feature of a photonic CES is that its group velocity points in only one direction, which is determined by the sign of the field that breaks the time-reversal symmetry and the resulting unusual topological properties of the bulk band structure. To detect the possible presence of non-trivial topological band properties in a photonic-crystal system it is sufficient^{5–7} to compute its Chern numbers. (Although the original proposal^{5,6} focused on ‘Dirac points’, it is not necessary to restrict to such band structures; thus, the use of a variety of photonic-crystal systems is possible⁷.) The Chern number of band n of a two-dimensional (2D) periodic photonic crystal is an integer defined by⁶

$$C_n = \frac{1}{2\pi i} \int_{\text{BZ}} d^2k \left(\frac{\partial A_y^{nn}}{\partial k_x} - \frac{\partial A_x^{nn}}{\partial k_y} \right)$$

where the k -space integral is performed over the first Brillouin zone and the Berry connection⁶ is given by

$$A^{nn'}(\mathbf{k}) \equiv i \langle \mathbf{E}_{n\mathbf{k}} | \nabla_{\mathbf{k}} | \mathbf{E}_{n'\mathbf{k}} \rangle = i \int d^2r \varepsilon(\mathbf{r}) \mathbf{E}_{n\mathbf{k}}^*(\mathbf{r}) \cdot [\nabla_{\mathbf{k}} \mathbf{E}_{n'\mathbf{k}}(\mathbf{r})]$$

where $\mathbf{E}_{n\mathbf{k}}$ is the periodic part of the electric-field Bloch function¹⁶, an asterisk denotes complex conjugation and $\varepsilon(\mathbf{r})$ denotes the dielectric function. Because the Chern number characterizes the winding number of the phase of the Bloch functions around the boundary of the first Brillouin zone¹¹, it is a ‘global’ or ‘topological’ property of the entire band and is very robust against structural perturbations¹⁰. Notably, it can be non-zero only if the system lacks time-reversal symmetry⁹.

One of the most interesting properties of QHE systems is that the Chern numbers have a direct physical significance: a finite crystal that supports bulk bands with non-zero Chern numbers also supports unidirectional CESs at its boundary at energies within bulk band gaps opened by the applied d.c. magnetic field. Moreover, the number of CESs turns out to be equal to the sum of the Chern numbers of all the bulk bands of lower energy¹³. Although this result has been formally proven only in a tight-binding QHE system, it is believed to be independent of the details of the underlying model, such as the structure of the lattice and the edge. Its validity in photonic-crystal systems was originally predicted in refs 5, 6, and corroborated through a formal mapping⁷ to a ‘zero-field QHE’ system¹² and *ab initio* numerical simulations of Maxwell’s equations⁷. It is important to emphasize that although CESs have so far been experimentally observed only in electronic (that is, fermionic) systems, the phenomenon should actually be independent of the underlying particle statistics because the Chern number is defined in terms of single-particle Bloch functions. An experimental verification would therefore provide strong support for the generalization of topological band theories and their applications to classical and bosonic systems.

The ability to work with photonic-crystal band structures without Dirac points has allowed us to design an experimentally viable photonic-crystal system⁷ for the observation of CESs. Our experimental system (Fig. 1) involves a gyromagnetic, 2D-periodic photonic crystal consisting of a square lattice of ferrite rods in air (details of the structure and materials used can be found in Methods), bounded on one side by a non-magnetic metallic cladding. The interface between the photonic crystal and the cladding acts as a confining edge or waveguide for CESs. (Without this cladding, the CESs at the air edges of the photonic crystal would simply radiate away.) Neglecting absorption losses and nonlinear effects, we would expect power transmission of a CES along this waveguide to be independent of the waveguide geometry and also immune to backscattering from disorder, obstacles and defects.

¹Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [†]Present address: Yale University, New Haven, Connecticut 06520, USA. *These authors contributed equally to this work.

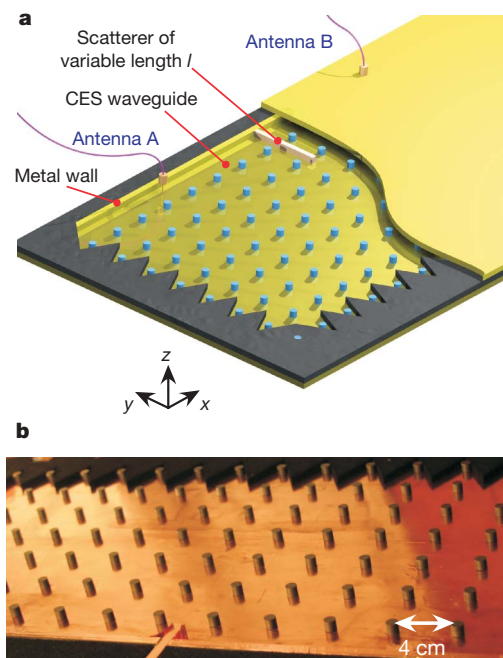


Figure 1 | Microwave waveguide supporting CESs. **a**, Schematic of the waveguide composed of an interface between a gyromagnetic photonic-crystal slab (blue rods) and a metal wall (yellow). The structure is sandwiched between two parallel copper plates (yellow) for confinement in the z direction and surrounded with microwave-absorbing foams (grey regions). Two dipole antennas, A and B, serve as feeds and/or probes. A variable-length (l) metal obstacle (orange) with a height equal to that of the waveguide (7.0 mm) is inserted between the antennas to study scattering. A 0.20-T d.c. magnetic field is applied along the z direction using an electromagnet (not shown). **b**, Top view (photograph) of the actual waveguide with the top plate removed.

Before we discuss the results of our measurements, we will first describe how we arrived at this particular choice of experimental system. We chose rods in air for the basic photonic-crystal geometry because of ease of fabrication. We then performed a series of numerical simulations for a variety of rod sizes and lattice constants on a model 2D photonic-crystal system to optimize the band structure and compute corresponding band Chern numbers using material parameters appropriate to a low-loss ferrite (Methods). Our numerical simulations predicted that when the ferrite rods in this photonic crystal are magnetized to manifest gyrotropic permeability (which breaks time-reversal symmetry), a gap opens between the second and third transverse magnetic (TM) bands. Moreover, the second, third and fourth bands of this photonic crystal acquire Chern numbers of 1, -2 and 1, respectively. This result follows from the C_{4v} symmetry of a non-magnetized crystal¹⁷. The results of our simulations for the photonic crystal with metallic cladding are presented in Fig. 2. (Similar numerical results were obtained in ref. 7, albeit using a different material system and geometry.) Here we show the calculated field patterns of a photonic CES residing in the second TM band gap (between the second and the third bands). Because the sum of the Chern numbers over the first and second bands is 1, exactly one CES is predicted to exist at the interface between the photonic crystal and the metal cladding. The simulations clearly predict that this photonic CES is unidirectional. As side-scattering is prohibited by the bulk photonic band gaps in the photonic crystal and in the metallic cladding, the existence of the CES forces the feed dipole antennas (which would radiate omnidirectionally in a homogeneous medium) to radiate only towards the right (Fig. 2a, c). Moreover, the lack of any backwards-propagating mode eliminates the possibility of backscattering, meaning that the fields can continuously navigate around obstacles, as shown in Fig. 2b. Hence, the scattering from the

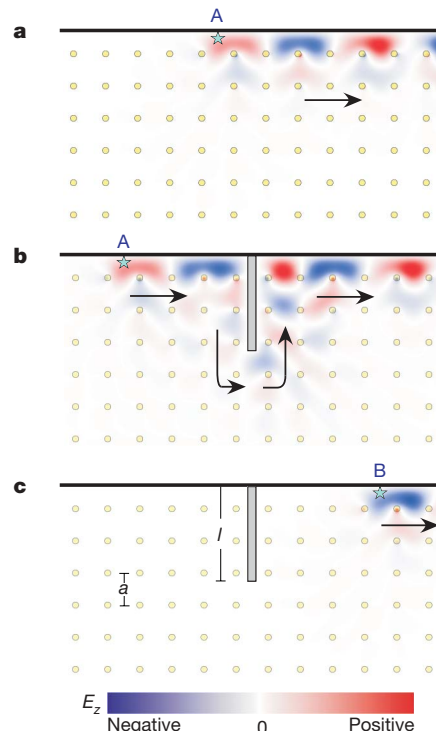


Figure 2 | Photonic CESs and effects of a large scatterer. **a**, CES field distribution (E_z) at 4.5 GHz in the absence of the scatterer, calculated from finite-element steady-state analysis (COMSOL Multiphysics). The feed antenna (star), which is omnidirectional in homogeneous media (Supplementary Information), radiates only to the right along the CES waveguide. The black arrow represents the direction of the power flow. **b**, When a large obstacle (three lattice constants long) is inserted, forward transmission remains unchanged because backscattering and side-scattering are entirely suppressed. The calculated field pattern (colour scale) illustrates how the CES wraps around the scatterer. **c**, When antenna B is used as feed antenna, negligible power is transmitted to the left, as the backwards-propagating modes are evanescent. a , lattice constant.

obstacle results only in a change of the phase (compare Fig. 2a and Fig. 2b) of the transmitted radiation, with no reduction in amplitude.

For CESs to be readily measurable in the laboratory (where it is necessary to use a photonic crystal of finite and manageable size) they must be spatially well localized, and this requires the photonic band gaps containing the states to be large. The sizes of the band gaps that contain CESs (and the frequencies at which they occur) are determined by the gyromagnetic constants of the ferrite rods constituting the photonic crystal. Under a d.c. magnetic field, microwave ferrites exhibit a ferromagnetic resonance at a frequency determined by the strength of the applied field¹⁸. Near this frequency, the Voigt parameter, $V = |\mu_{xy}|/|\mu_{xx}|$ (where μ_{xx} and μ_{xy} are diagonal and off-diagonal elements of the permeability tensor, respectively), which is a direct measure of the strength of the gyromagnetic effect, is of order one. Such ferromagnetic resonances are among the strongest low-loss gyrotropic effects at room temperature and subtesla magnetic fields. Using ferrite rods composed of vanadium-doped calcium-iron-garnet under a biasing magnetic field of 0.20 T (Methods and Supplementary Information), we achieved a relative bandwidth of 6% for the second TM band gap (around 4.5 GHz in Fig. 3b). As discussed earlier, this is the gap predicted to support a CES at the interface of the photonic crystal with the metallic wall. We emphasize again that band gaps with trivial topological properties (that is, for which the Chern numbers of the bulk bands of lower frequencies sum to zero), such as the first TM band gap (around 3 GHz in Fig. 3b), do not support CESs. All of the insight gained from the model 2D photonic-crystal system was then incorporated into the final design (Fig. 1). To emulate the states of the 2D photonic crystal, the final design

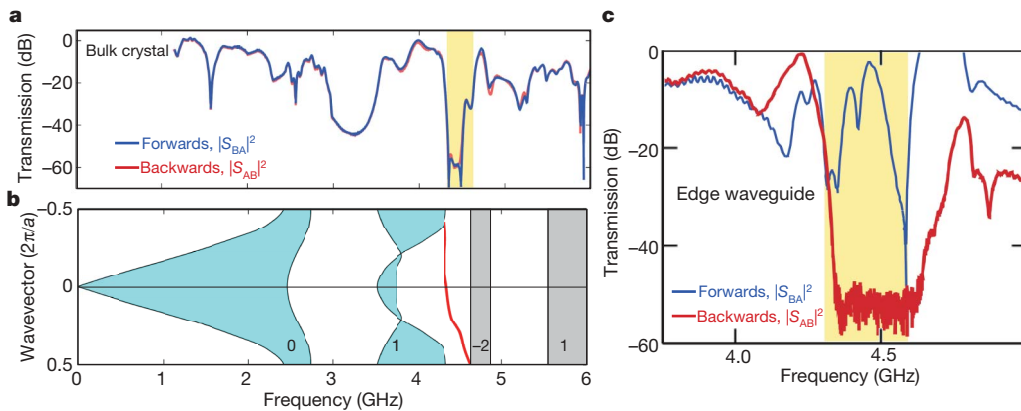


Figure 3 | CES-facilitated waveguiding in a photonic crystal. **a**, Forward and backward transmission spectra measured using only the bulk photonic crystal in the set-up shown in Fig. 1 (that is, without the metal cladding and obstacle), with the antennas placed in the interior of the photonic crystal, in a 0.20-T d.c. magnetic field. The bulk transmission is reciprocal, with photonic band gaps at 3.3 and 4.5 GHz. **b**, Calculated projected photonic-crystal band structure (blue and grey areas). Included is the CES (red curve) that exists at the interface between the metal cladding and the photonic

crystal. The grey areas are bulk bands with ill-defined band-edges due to large absorption near the ferromagnetic resonance. Each band's Chern number is labelled. **c**, Measured transmission spectra upon inclusion of the metal cladding and antennas placed as shown in Fig. 1. In the resulting CES waveguide, there is very high contrast between the forward and backward transmissions for frequencies in the second band gap (yellow), around 4.5 GHz. This striking unidirectionality indicates the existence of a CES.

involved fabrication of a three-dimensional (3D) photonic-crystal slab structure equivalent to the model 2D photonic-crystal system, made from gyromagnetic rods with parallel metallic plates on the top and bottom, spaced to support only transverse electromagnetic modes (identical to the TM modes in the 2D photonic crystal; see Methods). A copper wall was then added at the edge of the photonic-crystal slab to provide the required cladding.

In our experiments, the band gaps and the CES waveguide were characterized using two-port vector network analysis using a pair of dipole antennas, labelled A and B in Fig. 1a (Methods). First, to characterize the band gap, we inserted antennas A and B into the interior of the photonic crystal far from the edges and eight lattice constants apart. We observed the second band gap with a 50-dB extinction for both forward and backward transmission (with respective transmission coefficients S_{BA} and S_{AB} ; Fig. 3a). The frequency ranges of both the first and the second band gaps agree well with our predicted band structure calculations (no adjustable parameters; Fig. 3b). Next, to characterize the CESs, we measured the transmission spectra with the apparatus as illustrated in Fig. 1a (Methods). At frequencies within the second band gap, we observed a strong forward transmission, approximately 50 dB greater than the backward transmission at mid-gap frequencies (Fig. 3c). Over much of this frequency range, the backward transmission was below the noise floor of the network analyser, which suggests an even greater actual contrast. This difference of more than five orders of magnitude in power transmission, over a distance of only eight lattice constants, confirms that backwards-propagating modes are highly evanescent, as predicted.

We tested the robustness of the unidirectional propagation by studying the effect of a large obstacle on transmission. We gradually inserted a conducting barrier across the waveguide, blocking the direct path between antennas A and B. The measured transmission behaviour at different stages of the insertion (Fig. 4) remains basically the same as that in Fig. 3c: the transmission between 4.35 and 4.62 GHz remains strongly non-reciprocal, with a 40–50-dB difference between the forward and backward transmissions. This finding agrees with the theoretical prediction that power transmission by means of CESs is fundamentally insensitive to scattering from arbitrarily large defects (Fig. 2b). This behaviour is a distinguishing feature of the present waveguide. In a conventional waveguide, insertion of such a large obstacle would cause very large backscattering and significantly reduced transmission to the output. For example, in a photonic crystal constructed using ordinary dielectric rods and with identical

dimensions (Supplementary Information), a similar barrier length of 1.65 lattice constants reduces forward transmission by four orders of magnitude. This measurement further confirms that the backwards-propagating modes are purely evanescent, and not merely lossy. If lossy backwards-propagating modes existed in the system, a large defect would scatter a significant portion of energy into them, essentially converting backscattering into loss. The forward transmission in the presence of the large defect would be much smaller than in

dimensions (Supplementary Information), a similar barrier length of 1.65 lattice constants reduces forward transmission by four orders of magnitude. This measurement further confirms that the backwards-propagating modes are purely evanescent, and not merely lossy. If lossy backwards-propagating modes existed in the system, a large defect would scatter a significant portion of energy into them, essentially converting backscattering into loss. The forward transmission in the presence of the large defect would be much smaller than in

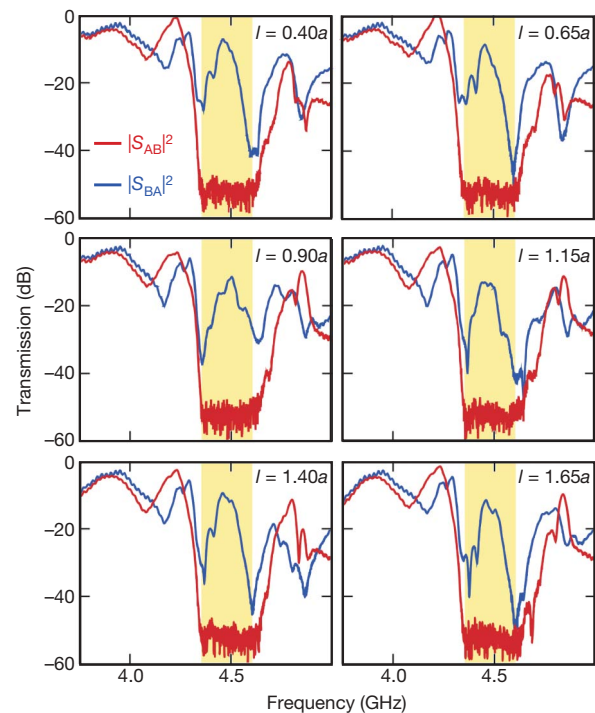


Figure 4 | CES transmission spectra in the presence of a large scatterer. The length of the obstacle, l , was gradually varied from $0.40a$ to $1.65a$ (lattice constant, $a = 40$ nm); this induced only minor differences in the forward transmission near the mid-gap frequency of 4.5 GHz. The lack of any significant changes in the forward transmission, and non-reciprocity ($|S_{AB}| \ll |S_{BA}|$) with large increases in the size of the scatterer, indicate that the CES can travel around the obstacle without scattering or reflections, as predicted by simulations. The experimental parameters remained unchanged from the measurement in Fig. 3c.

the defect-free case. Existing optical isolators, such as those relying on Faraday rotation or non-reciprocal phase shifts, absorb or radiate backwards-propagating light in this way. Thus, the unidirectional guiding of a CES is fundamentally different from how optical isolators operate.

The experimental establishment of topological photonic states opens a wide range of future opportunities. First, our realization of nontrivial topological Chern numbers in a classical photonic system raises the possibility of using photonic systems to realize other classes of topological quantum numbers that are of interest in condensed-matter physics. Examples include the Z_2 topological number associated with the quantum spin Hall effect^{19–22} and the ‘Hopf number’ in certain 3D insulators²³. Photonic crystals are attractive for such investigations because parameters such as lattice constants and unit-cell geometries can be chosen in a fully controlled manner¹⁶, unlike in most electronic systems. Second, the fact that the CESs in the present system are immune to scattering from disorder ensures that the design is tolerant of fabrication imperfections, such as variations in the lattice constant or the exact position of the guiding edge; this could enable implementation of extremely robust waveguides. Finally, photonic CESs might prove useful in applications involving isolators²⁴ or slow light^{25,26}. In conventional slow-light systems, disorder induces backscattering that increases quadratically with reduced group velocity²⁷, making them very sensitive to disorder. Although the experiments described here were conducted at gigahertz frequencies, this operating frequency can be increased simply by applying a stronger d.c. magnetic field¹⁸. Extension into the terahertz range might be achieved using metamaterials that resonantly enhance the magnetic activity^{28–30}. Further extension to the optical regime is challenging, given the losses and weak gyrotropic effects in currently known materials.

METHODS SUMMARY

The gyromagnetic photonic crystal was constructed using a square array (lattice constant, $a = 40$ mm) of vanadium-doped calcium-iron-garnet (VCIG; TCI ceramics NG-1850) rods. Balancing the need for a large Voigt parameter against the drawback of absorption loss in the vicinity of the ferromagnetic resonance (5.6 GHz), we designed the rod radius to be 3.9 mm and a to be 40 mm to maximize the bandwidth of the band gap without suffering excessive loss. A 16×10 array was used to measure the band gap of a bulk crystal and a 16×7 array was used to study the waveguide and the effect of scattering. The VCIG ferrite has a measured relative permittivity of $\epsilon_r = 14.63$ and a loss tangent of $\tan \delta = 0.00010$. The saturation magnetization was measured to be $M_s = 1.52 \times 10^5 \text{ A m}^{-1}$, with a 3-dB linewidth of the ferromagnetic resonance at $\Delta H = 1.03 \times 10^3 \text{ A m}^{-1}$. Using the cyclotron electromagnet at Massachusetts Institute of Technology, we applied a d.c. magnetic field of 0.20 T along the out-of-plane z direction, with a spatial non-uniformity of less than 1.5%. The d.c. magnetic field breaks the time-reversal symmetry in the photonic crystal. The magnetic field strength was measured and calibrated using a LakeShore Model 410 gaussmeter.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 1 June; accepted 15 July 2009.

- von Klitzing, K., Dorda, G. & Pepper, M. New method for high-accuracy determination of the fine-structure constant based on quantized Hall resistance. *Phys. Rev. Lett.* **45**, 494–497 (1980).
- Tsui, D. C., Stormer, H. L. & Gossard, A. C. Two-dimensional magnetotransport in the extreme quantum limit. *Phys. Rev. Lett.* **48**, 1559–1562 (1982).
- Novoselov, K. S. *et al.* Two-dimensional gas of massless Dirac fermions in graphene. *Nature* **438**, 197–200 (2005).
- Zhang, Y. B., Tan, Y. W., Stormer, H. L. & Kim, P. Experimental observation of the quantum Hall effect and Berry’s phase in graphene. *Nature* **438**, 201–204 (2005).

- Haldane, F. D. M. & Raghu, S. Possible realization of directional optical waveguides in photonic crystals with broken time-reversal symmetry. *Phys. Rev. Lett.* **100**, 013904 (2008).
- Raghu, S. & Haldane, F. D. M. Analogs of quantum-Hall-effect edge states in photonic crystals. *Phys. Rev. A* **78**, 033834 (2008).
- Wang, Z., Chong, Y. D., Joannopoulos, J. D. & Soljacic, M. Reflection-free one-way edge modes in a gyromagnetic photonic crystal. *Phys. Rev. Lett.* **100**, 013905 (2008).
- Prange, R. E. & Girvin, S. M. (eds). *The Quantum Hall effect* (Springer, 1987).
- Thouless, D. J., Kohmoto, M., Nightingale, M. P. & Denny, M. Quantized hall conductance in a two-dimensional periodic potential. *Phys. Rev. Lett.* **49**, 405–408 (1982).
- Simon, B. Holonomy, the quantum adiabatic theorem, and Berry phase. *Phys. Rev. Lett.* **51**, 2167–2170 (1983).
- Kohmoto, M. Topological invariant and the quantization of the Hall conductance. *Ann. Phys.* **160**, 343–354 (1985).
- Haldane, F. D. M. Model for a quantum Hall effect without Landau levels: condensed-matter realization of the ‘parity anomaly’. *Phys. Rev. Lett.* **61**, 2015–2018 (1988).
- Hatsugai, Y. Chern number and edge states in the integer quantum Hall effect. *Phys. Rev. Lett.* **71**, 3697–3700 (1993).
- Yablonovitch, E. Inhibited spontaneous emission in solid-state physics and electronics. *Phys. Rev. Lett.* **58**, 2059–2062 (1987).
- John, S. Strong localization of photons in certain disordered dielectric superlattices. *Phys. Rev. Lett.* **58**, 2486–2489 (1987).
- Joannopoulos, J. D., Johnson, S. G., Winn, J. N. & Meade, R. D. *Photonic Crystals: Molding the Flow of Light* (Princeton Univ. Press, 2008).
- Chong, Y. D., Wen, X. G. & Soljacic, M. Effective theory of quadratic degeneracies. *Phys. Rev. B* **77**, 235125 (2008).
- Pozar, D. M. *Microwave Engineering* 2nd edn (Wiley, 1998).
- Murakami, S., Nagaosa, N. & Zhang, S.-C. Dissipationless quantum spin current at room temperature. *Science* **301**, 1348–1351 (2003).
- Kane, C. L., Mele, E. J. & Z. (2) topological order and the quantum spin Hall effect. *Phys. Rev. Lett.* **95**, 146802 (2005).
- Bernevig, B. A., Hughes, T. L. & Zhang, S. C. Quantum spin Hall effect and topological phase transition in HgTe quantum wells. *Science* **314**, 1757–1761 (2006).
- Hsieh, D. *et al.* A topological Dirac insulator in a quantum spin Hall phase. *Nature* **452**, 970–975 (2008).
- Moore, J. E., Ran, Y. & Wen, X.-G. Topological surface states in three-dimensional magnetic insulators. *Phys. Rev. Lett.* **101**, 186805 (2008).
- Yu, Z. F. & Fan, S. H. Complete optical isolation created by indirect interband photonic transitions. *Nature Photon.* **3**, 91–94 (2009).
- Baba, T. Slow light in photonic crystals. *Nature Photon.* **2**, 465–473 (2008).
- Thevenaz, L. Slow and fast light in optical fibres. *Nature Photon.* **2**, 474–481 (2008).
- Povinelli, M. L. *et al.* Effect of a photonic band gap on scattering from waveguide disorder. *Appl. Phys. Lett.* **84**, 3639–3641 (2004).
- Pendry, J. B., Holden, A. J., Robbins, D. J. & Stewart, W. J. Magnetism from conductors and enhanced nonlinear phenomena. *IEEE Trans. Microw. Theory Tech.* **47**, 2075–2084 (1999).
- Yen, T. J. *et al.* Terahertz magnetic response from artificial materials. *Science* **303**, 1494–1496 (2004).
- Linden, S. *et al.* Magnetic response of metamaterials at 100 terahertz. *Science* **306**, 1351–1353 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are very grateful to P. Fisher and U. J. Becker for generously providing access to the synchrotron magnet at Massachusetts Institute of Technology. We should like to thank I. Chuang, P. Bermel, J. Bravo-Abad, S. Johnson and P. Rakich for comments. This work was supported in part by the Materials Research Science and Engineering Program of the US National Science Foundation under award number DMR-0819762, and also in part by the US Army Research Office through the Institute for Soldier Nanotechnologies under contract no. W911NF-07-D-0004.

Author Contributions Z.W., Y.C., J.D.J. and M.S. designed the photonic-crystal system, analysed the data and wrote the manuscript. Z.W. and Y.C. fabricated the structure and performed all the experimental measurements.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to Z.W. (zhwang@mit.edu).

METHODS

Parallel-plate waveguide for out-of-plane confinement. The unidirectional CES waveguide was designed to reproduce the dispersion relation and the modal profile of a topological edge mode of a 2D gyromagnetic photonic crystal, using a 3D structure with a finite height. The out-of-plane confinement in the z direction was achieved using two parallel horizontal copper plates, separated by 7.0 mm. This structure is known as a parallel-plate waveguide in microwave engineering¹⁸. It supports TEM modes with electric fields pointing in the out-of-plane z direction and magnetic fields parallel to the x - y plane. This polarization is identical to the TM modes in 2D photonic crystals where topological modes have been proposed to exist⁷. Between the two plates, the electromagnetic fields of TEM modes are also uniform along the z direction, as in a 2D system. This 3D structure therefore closely mimics a 2D system and is considered to be quasi-2D. When operated below 21 GHz, the waveguide supports only TEM modes.

Single-mode microwave CES waveguide and absorbing boundaries. Similar to the case of conventional waveguides, if the edge waveguide has too large a cross-sectional area it could lead to multimode operation, causing both a unidirectional CES as well as conventional bidirectional modes to be present in the waveguide. To ensure that only a CES is present in the measurement set-up, we chose the distance between the photonic crystal and the conducting copper wall to be 25 mm, which is narrow enough to eliminate all bidirectional modes at the frequencies of the second band gap. With a 6% relative bandwidth for this band gap, a CES is confined within three lattice constants of the edge, even around a large scatterer. The copper scatterer had a height of 7.0 mm and a width of 7.2 mm, with its maximum length mainly limited by the finite size of the crystal used in this experiment. Microwave-absorbing foam pieces were placed along the other three edges of the photonic crystal, to prevent the CES from circulating all the way around the boundary of the crystal. In addition, these foam pieces shielded the system from external interference.

Microwave transmission measurement for bulk crystals and for CESs. Two identically constructed antennas were inserted through the top copper plate, extending to contact the bottom copper plate. These antennas, labelled A and B in Fig. 1a, were connected by coaxial cables to the two ports of a Hewlett Packard 8719C vector network analyser, which measures the transmission coefficients S_{AB} and S_{BA} . Two-port short–open–load–through calibrations were performed at the coaxial adaptor. Therefore, measured S parameters contain a frequency-dependent insertion loss from the impedance mismatch between the antenna, the feed coaxial cable and the photonic-crystal waveguide, and from the transition between the balanced parallel plates and the unbalanced coax cable. This loss is reciprocal and does not affect the ratio of the transmission coefficients, $|S_{AB}/S_{BA}|$. Therefore, any substantial difference between $|S_{AB}|$ and $|S_{BA}|$ is an experimental signature of the unidirectionality of CESs. We extracted the forward and backward transmission spectra in a frequency sweep from 1 to 6 GHz. Each measurement was performed with an intermediate frequency of 20 Hz and four averages, with the power level normalized to the level at the band edges. To measure bulk band gaps (Fig. 3a), antennas A and B were located along the long axis of a 16×10 photonic crystal, eight lattice constants apart (Supplementary Information). For the CES waveguide (Figs 3c and 4), we performed the measurement with the feed and probe antennas located between the copper wall and the 16×7 photonic crystal, also eight lattice constants apart (Fig. 1a), and with the metal wall 9 mm away from each antenna.

Effects of material absorption loss. Most of the propagation loss in the present system may be attributed to two sources: the radiation losses originating from the finite width of the photonic-crystal cladding and the intrinsic material absorption associated with the ferromagnetic resonance. The radiation loss could be further reduced simply by increasing the number of unit cells in the lateral direction, whereas the absorption loss could in principle be further reduced by using monocrystalline yttrium–iron–garnet as the ferrite material¹⁸. The resultant attenuation length would be on the order of hundreds of lattice constants.

LETTERS

Early Palaeogene temperature evolution of the southwest Pacific Ocean

Peter K. Bijl¹, Stefan Schouten³, Appy Sluijs¹, Gert-Jan Reichert², James C. Zachos⁴ & Henk Brinkhuis¹

Relative to the present day, meridional temperature gradients in the Early Eocene age (~56–53 Myr ago) were unusually low, with slightly warmer equatorial regions¹ but with much warmer subtropical Arctic² and mid-latitude³ climates. By the end of the Eocene epoch (~34 Myr ago), the first major Antarctic ice sheets had appeared^{4,5}, suggesting that major cooling had taken place. Yet the global transition into this icehouse climate remains poorly constrained, as only a few temperature records are available portraying the Cenozoic climatic evolution of the high southern latitudes. Here we present a uniquely continuous and chronostratigraphically well-calibrated TEX₈₆ record of sea surface temperature (SST) from an ocean sediment core in the East Tasman Plateau (palaeolatitude ~65° S). We show that southwest Pacific SSTs rose above present-day tropical values (to ~34 °C) during the Early Eocene age (~53 Myr ago) and had gradually decreased to about 21 °C by the early Late Eocene age (~36 Myr ago). Our results imply that there was almost no latitudinal SST gradient between subequatorial and subpolar regions during the Early Eocene age (55–50 Myr ago). Thereafter, the latitudinal gradient markedly increased. In theory, if Eocene cooling was largely driven by a decrease in atmospheric greenhouse gas concentration⁶, additional processes are required to explain the relative stability of tropical SSTs given that there was more significant cooling at higher latitudes.

The Palaeogene temperature evolution of the Antarctic margin, particularly the Pacific sector, is still poorly resolved. One difficulty with obtaining relevant records close to the Antarctic continent is the general absence of biogenic carbonate in most marine facies, which hampers traditional $\delta^{18}\text{O}$ and/or Mg/Ca-based reconstructions of the subpolar temperature evolution. In the absence of biogenic carbonates, organic sea-surface-temperature proxies such as the tetraether index of lipids consisting of 86 carbon atoms (TEX₈₆)⁷ and the alkenone unsaturation index (U^{k}_{37})⁸ are required for reconstructing high-latitude climatic evolution^{2,9}.

We apply TEX₈₆ and U^{k}_{37} on a stratigraphically continuous sedimentary section from the southwest Pacific Ocean, drilled by the Ocean Drilling Program (ODP Leg 189 Site 1172, palaeolatitude ~65° S (ref. 10); Fig. 1). A full methodological description is available in Supplementary Information. The record contains an expanded succession of marginal marine sediments from the lower Palaeocene epoch to the upper Eocene (64–36 Myr ago), with tight chronostratigraphic control, including magnetostratigraphy¹¹ (Supplementary Fig. 2). The presence of typical trans-Antarctic organic-walled dinoflagellate cysts in the Tasman region indicates an Antarctic-derived northward-flowing Tasman Current throughout the Palaeogene, which is verified by experiments based on general circulation models¹² (Fig. 1). This Antarctic influence at the East Tasman Plateau (ETP) persisted until at least the early Late Eocene (~35.5 Myr ago), when

deepening of the Tasmanian Gateway lead to a reorganization of the Tasman and proto-Leeuwin ocean currents¹³.

According to the oldest part of the record, TEX₈₆-derived SSTs at the ETP gradually decreased from ~25 °C around 63 Myr ago to a minimum of ~20 °C around 58 Myr ago (Fig. 2a). During the Late Palaeocene and Early Eocene, Tasman SSTs gradually rose to tropical values of ~34 °C during the Early Eocene climatic optimum (EECO)⁶, between 53 and 49 Myr ago (Fig. 2a). A gradual cooling trend throughout the Middle Eocene (starting at the termination of the EECO ~49 Myr ago) arrived at temperatures of ~23 °C ~42 Myr ago, which is still relatively warm. Subsequently, an interruption of the cooling trend occurred at the Middle Eocene climatic optimum (MECO; ~40 Myr ago)¹⁴, followed by a relatively rapid SST decrease to ~21 °C in the early Late Eocene (Fig. 2a). The late Middle and Late Eocene TEX₈₆-based SSTs are supported by U^{k}_{37} SST estimates derived from the same samples (Supplementary Fig. 3). Both SST estimates also compare well with those for other Late Eocene (Southern Ocean) sites⁹. Unfortunately, sediments from the ETP older than the MECO did not contain alkenones for U^{k}_{37} SST reconstructions.

The Middle Eocene SSTs correspond closely to those from sections in New Zealand^{15,16}, according to records based on TEX₈₆ (Fig. 2a), Mg/Ca and $\delta^{18}\text{O}$, indicating regional consistency of our reconstructed SSTs. Also, trends in our Tasman SST record are remarkably similar to those in the global stack of benthic foraminiferal oxygen isotopes⁶ (Fig. 2b), which we updated and augmented with recently published

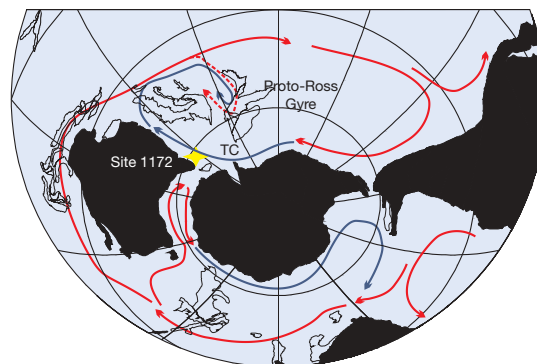


Figure 1 | Site location and surface currents. Palaeogeographic reconstruction for the South Pacific Ocean at Early–Middle Eocene times. Surface circulation¹² indicates the Antarctic-derived Tasman Current (TC) over the East Tasman Plateau. Palaeogeographic charts obtained from the Ocean Drilling Stratigraphic Network (ODSN); after ref. 26. The dashed red arrow around New Zealand indicates potential mixing of low-latitude surface waters (from the East Australian Current) with the TC.

¹Palaeoecology, Institute of Environmental Biology, Faculty of Science, Laboratory of Palaeobotany and Palynology, ²Department of Geochemistry, Faculty of Geosciences, Utrecht University, Budapestlaan 4, 3584 CD Utrecht, The Netherlands. ³Department of Marine Organic Biogeochemistry, NIOZ Royal Netherlands Institute of Sea Research, PO Box 59, 1790 AB Den Burg, Texel, The Netherlands. ⁴Earth and Planetary Sciences Department and Institute of Marine Sciences, University of California, Santa Cruz, 1156 High St, Santa Cruz, California 95064, USA.

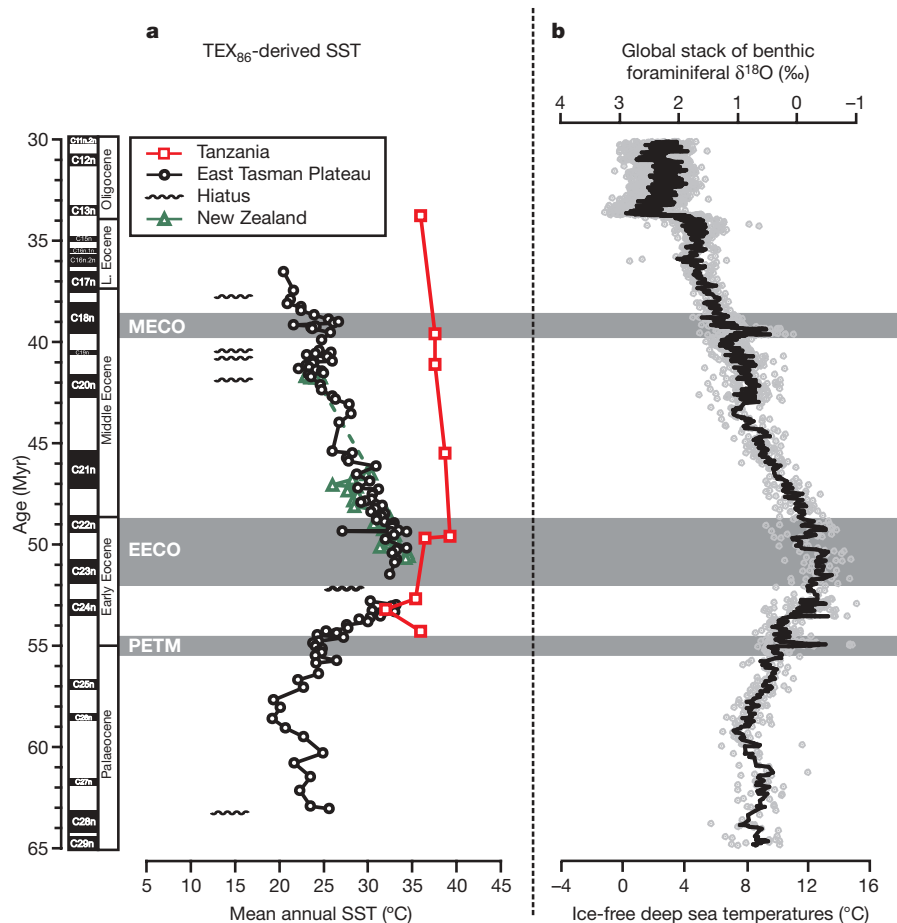


Figure 2 | Palaeogene deep-sea and sea surface temperatures. **a**, TEX₈₆ SST reconstructions from ODP Site 1172, New Zealand^{15,16} and Tanzania¹ (all according to the same calibration; see Supplementary Information). The black wiggly lines are short (~100 kyr)²⁷ and longer hiatuses at Site 1172. **b**, Global stack of benthic foraminiferal oxygen isotopes (grey data;

Supplementary Information). The temperature scale assumes ice-free conditions ($\delta^{18}\text{O}_{\text{SMOW}} = -1.2\text{‰}$, where $\delta^{18}\text{O}_{\text{SMOW}} = (^{18}\text{O}/^{16}\text{O})_{\text{sample}} / (^{18}\text{O}/^{16}\text{O})_{\text{SMOW}} - 1$; SMOW, standard mean ocean water), and indicates deep-sea temperatures. The black solid line reflects a five-point running average. PETM, Palaeocene–Eocene thermal maximum.

data (Supplementary Information). This correspondence between the two records (Supplementary Fig. 4) indicates that the regional SSTs co-varied with the SSTs where 'global' deep water was sourced. It has previously been suggested that the Southern Ocean was the main region of deep-water formation during the Palaeogene¹⁷.

In contrast, absolute SST estimates from the Tasman region are much higher than those inferred from the benthic foraminiferal oxygen isotopes (Fig. 2). Part of this discrepancy might be due to seasonality, with TEX₈₆ being slightly skewed towards summer temperatures and benthic foraminiferal $\delta^{18}\text{O}$ towards winter temperatures (Supplementary Information). Another possibility is that deep-water formation occurred in areas that were cooler than the Tasman sector. SST reconstructions based on bivalve-shell oxygen isotopes from Seymour Island on the Antarctic shelf, for example, yield much lower SSTs¹⁸. It is possible that the Antarctic margin was more susceptible to winter cooling than the open ocean, or that portions of the coast along the Southern Ocean gyres were somehow isolated from the southern edges of the Southern Ocean gyres. Another possibility is that the aragonite bivalve shells integrate temperature over a greater portion of the year. Regardless, the large SST difference between the Weddell Sea and the ETP would suggest a relatively steep gradient within a few degrees of latitude. Antarctica, being a polar continent, would most likely have experienced extremes in temperature, in particular having cool winters. Such conditions might have been recorded in the bivalves from the Weddell Sea but not in the more distal ETP. In turn, deep-water formation might have been restricted to the Antarctic shelf areas, such as the Weddell Sea.

Planktonic foraminiferal $\delta^{18}\text{O}$ analyses from equatorial regions previously indicated that Palaeogene low-latitude SSTs were the same, or even lower, than those of today¹⁹, a problem that puzzled palaeoclimate scientists for decades. The oxygen isotopic composition of planktonic foraminiferal tests in porous carbonate-rich pelagic facies were later found to be partially altered owing to recrystallization primarily during early diagenesis^{20,21}. In contrast, carbonate-poor and clay-rich facies typically found on the continental margins contain calcite shells without major diagenetic overprint¹. For the Eocene, such well-preserved planktonic foraminifera indicate near-equatorial SSTs that were greater than those of the present day, and agree with TEX₈₆-derived SSTs^{1,21}.

Another observation from well-preserved foraminifera and TEX₈₆ is that (sub)equatorial SSTs were remarkably stable throughout the Eocene¹ (Fig. 2). Stable low-latitude SSTs concomitant with high-latitude Eocene cooling thus suggests that there were increasing SST gradients during the Eocene. Although SST trends are often reconstructed using multi-proxy studies, the difference in absolute SSTs between various proxy reconstructions can be considerably large^{9,15,22}, even when measured on the same sediments. Despite the fact that multi-proxy approaches are generally encouraged in palaeoclimate studies, exclusion of such inter-proxy biases in latitudinal gradient reconstructions requires single-proxy SST records from around the world. Traditional calcite-based SST reconstructions are less suitable for this because calcite is only sparsely available in high-latitude sediments. The organic TEX₈₆ and U^k₃₇ SST proxies, however, can be used independently of latitude and are, hence, suitable for

single-proxy SST gradient reconstructions. Moreover, they do not require critical assumptions about ancient sea-water chemistry, unlike $\delta^{18}\text{O}$ and Mg/Ca .

We compiled Eocene TEX_{86} and U^{K}_{37} SST reconstructions from a suite of sedimentary records from localities worldwide and noted increased Middle Eocene latitudinal SST gradients in both hemispheres (Fig. 3), relative to the Early Eocene. These SST gradients are in general agreement with those found for terrestrial mean annual temperatures, based on Early–Middle Eocene fossil leaves²³. Adding the bivalve-based SST reconstructions from Seymour Island¹⁸ to our organic proxy data suggests a strong gradient between 60° and 70° S, which contrasts with the small gradient between 60° S and the Equator (Fig. 3). A part of this large Southern Ocean SST gradient might be due to biases between organic and calcite proxies. A large part, however, may realistically reflect the influence of the cool Antarctic interior, which cooled the Antarctic shelf. In contrast to the continental South Pole, the Arctic region is an oceanic basin. Instead of amplifying the seasonal cycle, the Arctic Ocean probably moderated seasonal extremes in the northern high-latitude greenhouse. Hence, Palaeogene latitudinal temperature gradients, like those of today, would have exhibited a high degree of asymmetry between the two hemispheres.

It has been suggested that the general warmth that characterized early Palaeogene climates was forced by high atmospheric greenhouse gas concentrations⁶. Concomitantly, the absence of polar ice sheets eliminated ice–albedo feedbacks in the Palaeogene greenhouse. The Middle–Late Eocene global cooling has been related to long-term atmospheric CO_2 decline, eventually resulting in the onset of major Antarctic glaciation around the Eocene/Oligocene boundary⁶. Our results imply that meridional temperature gradients markedly increased together with deep-sea cooling (Fig. 2)⁶. Although high latitudes cooled, tropical temperatures seem to have remained fairly stable throughout the Eocene (Fig. 3)¹. This observation raises questions concerning the precise role of decreasing atmospheric greenhouse gas concentrations in cooling the Eocene poles, as in theory²⁴ they should have cooled tropical regions as well. The role of potential

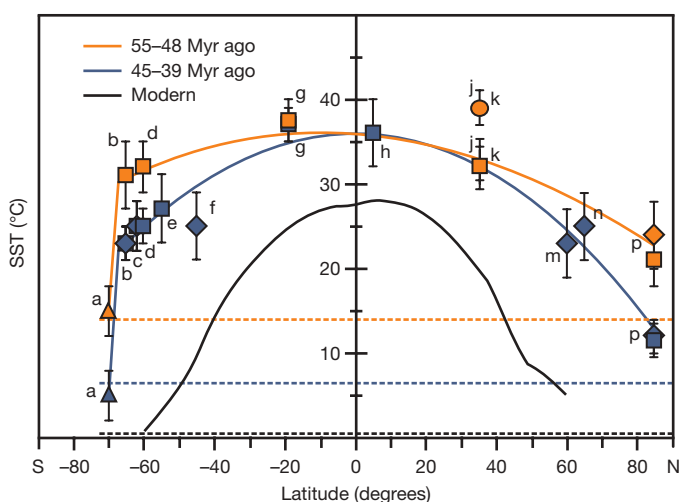


Figure 3 | Early and Middle Eocene latitudinal SST gradients. Bivalve-shell $\delta^{18}\text{O}$ (triangles), TEX_{86} (squares) and U^{K}_{37} (diamonds) SST reconstructions for the Early (orange) and mid-Middle (blue) Eocene. Data are from Seymour Island¹⁸ (a), the East Tasman Plateau (b), Deep Sea Drilling Project (DSDP) Site 277⁹ (c), New Zealand^{15,16} (d), DSDP Site 511⁹ (e), ODP Site 1090⁹ (f), Tanzania¹ (g), ODP Site 925⁹ (h), New Jersey³ (j, k; circle represents peak PETM SSTs³), ODP Site 336⁹ (m), ODP Site 913⁹ (n) and the Arctic Ocean^{2,28,29} (p) (Supplementary Fig. 1). Error bars indicate the range of variation. Gradients represent second-order polynomials, excluding bivalve-shell data. Black and dashed lines represent the present-day zonally averaged latitudinal temperature gradient³⁰ and age-specific deep-sea temperatures, respectively (Fig. 2b, ref. 6).

high-latitude climate feedbacks involving, for example, differences in cloud/water vapour distribution²⁵ might have been much more instrumental in the Middle Eocene climatic deterioration than previously thought. Another potential positive-feedback mechanism for high-latitude cooling would be ice–albedo feedback. However, the presence of substantial Middle Eocene continental ice is still equivocal given the general warmth and overall absence of conclusive physical evidence.

Received 31 March; accepted 6 August 2009.

- Pearson, P. N. *et al.* Stable tropical climate through the Eocene epoch. *Geology* **35**, 211–214 (2007).
- Sluijs, A. *et al.* Arctic late Paleocene–early Eocene paleoenvironments with special emphasis on the Paleocene–Eocene thermal maximum (Lomonosov Ridge, Integrated Ocean Drilling Program Expedition 302). *Paleoceanography* **23**, doi:10.1029/2007PA001495 (2008).
- Sluijs, A. *et al.* Environmental precursors to rapid light carbon injection at the Palaeocene/Eocene boundary. *Nature* **450**, 1218–1221 (2007).
- Zachos, J. C., Breza, J. R. & Wise, S. W. Jr. Early Oligocene ice sheet expansion on Antarctica: stable isotope and sedimentological evidence from Kerguelen Plateau, Southern Indian Ocean. *Geology* **20**, 569–573 (1992).
- Barker, P. F., Dieckmann, B. & Escutia, C. Onset of Cenozoic Antarctic glaciation. *Deep-Sea Res. II* **54**, 2293–2307 (2007).
- Zachos, J. C., Dickens, G. R. & Zeebe, R. E. An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. *Nature* **451**, 279–283 (2008).
- Schouten, S., Hopmans, E. C., Schefuß, E. & Sinninghe Damsté, J. S. Distributional variations in marine crenarchaeotal membrane lipids: a new tool for reconstructing ancient sea water temperatures? *Earth Planet. Sci. Lett.* **204**, 265–274 (2002).
- Müller, P. J., Kirst, G., Rohland, G., von Storch, I. & Rosell-Melé, A. Calibration of the alkenone paleotemperature index U^{K}_{37} based on core-tops from the eastern South Atlantic and the global ocean (60°N–60°S). *Geochim. Cosmochim. Acta* **62**, 1757–1772 (1998).
- Liu, Z. *et al.* Global cooling during the Eocene–Oligocene climate transition. *Science* **323**, 1187–1190 (2009).
- Exon, N., Kennett, J. P. & Malone, M. (eds) *The Cenozoic Southern Ocean: Tectonics, Sedimentation, and Climate Change between Australia and Antarctica* (Geophys. Monogr. Ser. 151, American Geophysical Union, 2004).
- Stickley, C. E. *et al.* in *Proc. Ocean Drilling Program, Scientific Results* (eds Exon, N. F., Kennett, J. P. & Malone, M. J.) 1–57 (2004).
- Huber, M. *et al.* Eocene circulation of the Southern Ocean: was Antarctica kept warm by subtropical waters? *Paleoceanography* **19**, doi:10.1029/2004PA001014 (2004).
- Stickley, C. E. *et al.* Timing and nature of the deepening of the Tasmanian Gateway. *Paleoceanography* **19**, doi:10.1029/2004PA001022 (2004).
- Bohaty, S. M. & Zachos, J. C. Significant Southern Ocean warming event in the late Middle Eocene. *Geology* **31**, 1017–1020 (2003).
- Hollis, C. J. *et al.* Tropical sea temperatures in the high latitude South Pacific during the Eocene. *Geology* **37**, 99–102 (2009).
- Burgess, C. E. *et al.* Middle Eocene climate cyclicity in the southern Pacific: implications for global ice volume. *Geology* **36**, 651–654 (2008).
- Thomas, D. J., Bralower, T. J. & Jones, C. E. Neodymium isotopic reconstruction of the Late Paleocene–Early Eocene thermohaline circulation. *Earth Planet. Sci. Lett.* **209**, 309–322 (2003).
- Ivany, L. C. *et al.* Eocene climate record of a high southern latitude continental shelf: Seymour Island, Antarctica. *Geol. Soc. Am. Bull.* **120**, 659–678 (2008).
- Barron, E. J. Eocene equator-to-pole surface ocean temperatures: a significant climate problem? *Paleoceanography* **2**, 729–739 (1987).
- Schrag, D. P., DePaolo, D. J. & Richter, F. M. Reconstructing past sea surface temperatures: correcting for diagenesis of bulk marine carbonate. *Geochim. Cosmochim. Acta* **59**, 2265–2278 (1995).
- Pearson, P. N. *et al.* Warm tropical sea surface temperatures in the Late Cretaceous and Eocene epochs. *Nature* **413**, 481–487 (2001).
- Huber, M. A hotter greenhouse? *Science* **321**, 353–354 (2008).
- Greenwood, D. R. & Wing, S. L. Eocene continental climates and latitudinal temperature gradients. *Geology* **23**, 1044–1048 (1995).
- Huber, M. & Sloan, L. C. Heat transport, deep waters, and thermal gradients: coupled simulation of an Eocene ‘greenhouse’ climate. *Geophys. Res. Lett.* **28**, 3481–3484 (2001).
- Abbot, D. S. & Tziperman, E. Sea ice, high-latitude convection, and equable climates. *Geophys. Res. Lett.* **35**, doi:10.1029/2007GL032286 (2008).
- Hay, W. W. *et al.* Alternative global Cretaceous paleogeography. *Spec. Pap. Geol. Soc. Am.* **332**, 1–47 (1999).
- Röhl, U. *et al.* in *The Cenozoic Southern Ocean: Tectonics, Sedimentation, and Climate Change between Australia and Antarctica* (eds Exon, N., Kennett, J. P. & Malone, M.) 127–151 (Geophys. Monogr. Ser. 151, American Geophysical Union, 2004).
- Sangiorgi, F. *et al.* Cyclicity in the middle Eocene central Arctic Ocean sediment record: orbital forcing and environmental response. *Paleoceanography* **23**, doi:10.1029/2007PA001487 (2008).

29. Weller, P. & Stein, R. Paleogene biomarker records from the central Arctic Ocean (Integrated Ocean Drilling Program Expedition 302): organic carbon sources, anoxia, and sea surface temperature. *Paleoceanography* **23**, doi:10.1029/2007PA001472 (2008).
30. Shea, D. J., Trenberth, K. E. & Reynolds, R. W. A global monthly sea surface temperature climatology. *J. Clim.* **5**, 987–1001 (1992).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Funding for this research was provided by Utrecht University, the Netherlands Organisation for Scientific Research (VICI grant to S.S.; VENI grant to A.S.) and the LPP Foundation. This research used samples and data provided by the Ocean Drilling Program (ODP). The ODP was sponsored by the US National Science

Foundation and participating countries under the management of Joint Oceanographic Institutions, Inc. G. Nobbe, E. van Bentum, E. Speelman, J. Ossebaar, A. Mets and E. Hopmans are thanked for technical support. We acknowledge C. J. Hollis, P. N. Pearson and P. F. Sexton for providing published data. A. J. P. (Sander) Houben, P. N. Pearson and M. Huber are thanked for critical comments.

Author Contributions P.K.B., S.S., H.B. and A.S. designed the research, P.K.B. and S.S. performed the organic geochemical analyses; P.K.B. updated the age model for ODP Site 1172 and performed the data compilations. All authors contributed to interpreting the data and writing the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to P.K.B. (p.k.bijl@uu.nl).

LETTERS

Rapid ascent of rhyolitic magma at Chaitén volcano, Chile

Jonathan M. Castro¹ & Donald B. Dingwell²

Rhyolite magma has fuelled some of the Earth's largest explosive volcanic eruptions¹. Our understanding of these events is incomplete, however, owing to the previous lack of directly observed eruptions. Chaitén volcano, in Chile's northern Patagonia, erupted rhyolite magma unexpectedly and explosively on 1 May 2008 (ref. 2). Chaitén residents felt earthquakes about 24 hours before ash fell in their town and the eruption escalated into a Plinian column. Although such brief seismic forewarning of a major explosive basaltic eruption has been documented³, it is unprecedented for silicic magmas. As precursory volcanic unrest relates to magma migration from the storage region to the surface, the very short pre-eruptive warning at Chaitén probably reflects very rapid magma ascent through the sub-volcanic system. Here we present petrological and experimental data that indicate that the hydrous rhyolite magma at Chaitén ascended very rapidly, with velocities of the order of one metre per second. Such rapid ascent implies a transit time from storage depths greater than five kilometres to the near surface in about four hours. This result has implications for hazard mitigation because the rapidity of ascending rhyolite means that future eruptions may provide little warning.

Geophysical precursors to volcanic eruptions, such as volcano-tectonic earthquakes, tremor and deformation, all reflect magma migration beneath the volcano as the magma develops an ascent path^{4,5}. Such signals are crucial for volcano monitoring, and increasingly, as the source mechanisms of seismicity are identified, eruption forecasting^{6–9}. A critical unknown that has limited the accuracy of eruption forecasting is the rate of magma rise before an explosive eruption: this parameter controls not only degassing behaviour and flow rheology^{10,11}, but also the timescale of accompanying precursory unrest and pre-eruptive warning¹². A vast majority of andesite and dacite volcanic eruptions were preceded by weeks to months of precursory unrest, consistent with long magma ascent times and correspondingly sluggish (some centimetres per second) rise velocities¹³. This pattern was broken on 1 May 2008 when Chaitén volcano, Chile, erupted with almost no warning at all. This explosive rhyolite eruption, the first ever to be scientifically monitored², provides a unique opportunity to assess the conditions of pre-eruptive magma storage and ascent at rhyolite volcanoes. Of particular interest is the extreme suddenness of the eruption, because this implies that rhyolite is highly mobile in the shallow crust. Here we constrain the storage conditions and pre-eruptive ascent velocity of rhyolite magma at Chaitén by experimentally reproducing key mineralogical and textural characteristics of pumice erupted from the volcano.

Pre-eruptive unrest at Chaitén began on 30 April 2008 at about 20:00 h Chilean Local Time (CLT) when residents of Chaitén town, about 10 km southwest of the volcano, felt earthquakes strong enough to knock objects off shelves. They first observed ash fall in their town on 1 May 2008 at roughly 21:00 h (CLT). Seismic activity

continued through to 2 May 2008 when a large explosion and Plinian eruption column tore through a prehistoric obsidian dome in the Chaitén caldera. After a week of fluctuating Plinian and sub-Plinian activity, a new lava dome began to grow, and this activity is still continuing. The Plinian eruption plume distributed a broad swath of tephra throughout the Andes¹⁴. We collected samples of this ash blanket from two sites located about 10 km east-southeast of the vent, and at another about 2 km north of the vent. The tephra deposit comprises ash (~80% by volume), pumice lapilli and bombs (~17%), and obsidian fragments (~3%).

The pumice lapilli are rhyolitic in composition (Table 1) and nearly aphyric (<1 vol.% crystals). Crystals comprise both microphenocrysts (0.5–1.0 mm) and sparse microlites (<100 µm), which we identified as plagioclase and biotite in about 10% of the sampled ($n = 40$) pyroclasts. The microphenocryst mineral population comprises plagioclase, Fe-Ti oxides, orthopyroxene and biotite; however, many pumices are completely devoid of biotite.

Plagioclase compositions are relatively uniform (~An_{40–45}), aside from a few crystal cores as calcic as An₆₈. These microphenocrysts are invariably rounded with zoning patterns (Fig. 1b) characterized by jagged compositional boundaries, indicating several cycles of dissolution and growth¹⁵. The lack of euhedral overgrowth rims on these plagioclase microphenocrysts suggests that they were in a state of resorption before eruption. Orthopyroxene is euhedral, and has a restricted compositional range (En_{50–55}; Table 1). Fe-Ti oxides include both titanomagnetite and rare ilmenite. We did not obtain reliable compositional analyses of biotite owing to its small size (1–2 µm wide).

Plagioclase and orthopyroxene microphenocrysts contain abundant small (<50 µm) rhyolitic glass inclusions (Fig. 1; Table 1). These inclusions contain large vesicles (>20 µm) that could reflect the entrapment of volatile-saturated melt during crystal growth. The dominant volatile component of the glass is water (H₂O and OH[–]), and the concentrations (~1.3 to 2.3 wt%) could reflect pre and syn-eruptive degassing through cracks and cleavage planes (Supplementary Information). The glass inclusions contain no detectable CO₂ (detection limit ~10 p.p.m.).

Obsidian pyroclasts are mineralogically identical to the pumice; however, they contain larger plagioclase crystals (2–3 mm) and are generally more crystalline (~2–5 vol.%). The obsidians are low in H₂O (0.5–1 wt%; Supplementary Information) and devoid of CO₂. It is not possible to prove that the obsidian pyroclasts are juvenile or lithic fragments derived from the obsidian dome in the Chaitén caldera, because this prehistoric lava dome is compositionally indistinguishable from the new magma¹⁶. We therefore focus the rest of this analysis and discussion on the petrogenesis of the Chaitén pumice, which is undoubtedly a juvenile eruption product.

The crystal complement in the Chaitén rhyolite records key information about pre-eruptive magma storage and ascent, as

¹Centre National de la Recherche Scientifique (CNRS), Institut National des Sciences de l'Univers (INSU), Université d'Orléans, Université François Rabelais–Tours, Institut des Sciences de la Terre d'Orléans, UMR 6113, Campus Géosciences, 1A Rue de la Férollerie, 45071 Orléans cedex 2, France. ²Department of Earth and Environmental Sciences, Ludwig-Maximilians-Universität, 80333 München, Germany.

Table 1 | Representative compositions of the Chaitén pumice and microphenocryst mineral phases

Component	Matrix glass†	Bulk‡	Glass inclusions†	Plagioclase†	Magnetite	Ilmenite	Orthopyroxene†
Sample	Ch-1-08	Ch-1-08	Ch-inc-1	Ch-plg2-08	C1-grn1	C1-grn2	Ch-opx1-08
<i>n</i>	<i>n</i> = 76	<i>n</i> = 5	<i>n</i> = 10	<i>n</i> = 20	<i>n</i> = 9	<i>n</i> = 9	<i>n</i> = 15
SiO ₂	76.1 (0.5)	75.6 (0.4)	76.1 (0.3)	59.2 (0.3)	n.d.	n.d.	48.3 (0.5)
Al ₂ O ₃	13.7 (0.2)	13.9 (0.2)	13.0 (0.4)	26.1 (0.5)	2.57 (0.05)	0.17 (0.01)	1.69 (0.2)
TiO ₂	0.13 (0.01)	0.14 (0.03)	0.5 (0.1)	n.d.	8.7 (0.07)	45.6 (0.19)	0.13 (0.02)
Fe ₂ O ₃	1.27 (0.1)	1.5 (0.02)	1.42 (0.01)	0.16 (0.03)	90.3 (0.34)	54.6 (0.39)	31.2 (1.4)
MgO	0.28 (0.01)	0.26 (0.1)	0.29 (0.02)	n.d.	0.91 (0.03)	1.80 (0.04)	16.6 (0.3)
MnO	0.06 (0.01)	0.05 (0.01)	0.06 (0.03)	n.d.	0.54 (0.04)	0.86 (0.04)	1.79 (0.2)
Cr ₂ O ₃	n.d.	n.d.	n.d.	n.d.	0.02 (0.02)	0.02 (0.01)	n.d.
CaO	1.41 (0.01)	1.46 (0.02)	1.10 (0.03)	8.13 (0.2)	n.d.	n.d.	0.34 (0.03)
Na ₂ O	4.00 (0.1)	4.04 (0.02)	3.91 (0.2)	6.5 (0.06)	n.d.	n.d.	n.d.
K ₂ O	2.98 (0.04)	2.93 (0.05)	3.10 (0.1)	0.28 (0.02)	n.d.	n.d.	n.d.
P ₂ O ₅	0.04 (0.03)	0.06 (0.01)	0.01	n.d.	n.d.	n.d.	n.d.
SO ₂ (p.p.m.)	31 (15)	n.d.	500	n.d.	n.d.	n.d.	n.d.
Cl ⁻ (p.p.m.)	925 (66)	n.d.	3550	n.d.	n.d.	n.d.	n.d.
Total	100.0 (0.45)	99.9 (0.77)	99.5 (1.02)	100.4 (0.53)	103.0 (0.34)	103.0 (0.44)	100.1 (0.40)

† EPMA; ‡ Bulk pumice X-ray fluorescence analysis.

n.d.: not detected. An:Ab:Or = 58:40:02

characteristics of mineral phases are all functions of pressure (P), temperature (T) and the melt-H₂O content. These intensive parameters may be constrained through petrological experiments^{17,18}. One complication is that the plagioclase microphenocrysts appear to have been unstable in the melt before eruption, and as such could be xenocrystic. As we show below, with the exception of rare calcic cores, plagioclase compositions are compatible with the rhyolite melt over a range of $P_{\text{H}_2\text{O}}-T$ space. Furthermore, the abundance of glass inclusions in the microphenocrysts, whose major element compositions are identical to the pumice matrix glass, suggests that the plagioclase crystals are indeed primary.

We performed petrological experiments^{18–20} on a powdered pumice pyroclast over a range of $P_{\text{H}_2\text{O}}-T$ conditions (Supplementary Information). Given the explosive nature of the eruption, and the lack of CO₂ in melt inclusions, we assumed that the pre-eruption magma was water-saturated, and added just enough water to the pumice powder to achieve water-saturation. These experiments (Fig. 2) indicate that the crystal population is stable in hydrous rhyolite melt over a wide $P_{\text{H}_2\text{O}}-T$ range (~ 50 – 200 MPa; ~ 780 – 850 °C). Permissible magma storage conditions may be further bracketed by considering that the natural plagioclase ($\sim \text{An}_{40-45}$) and orthopyroxene ($\sim \text{En}_{50-55}$) are together reproduced at about $T < 825$ °C and

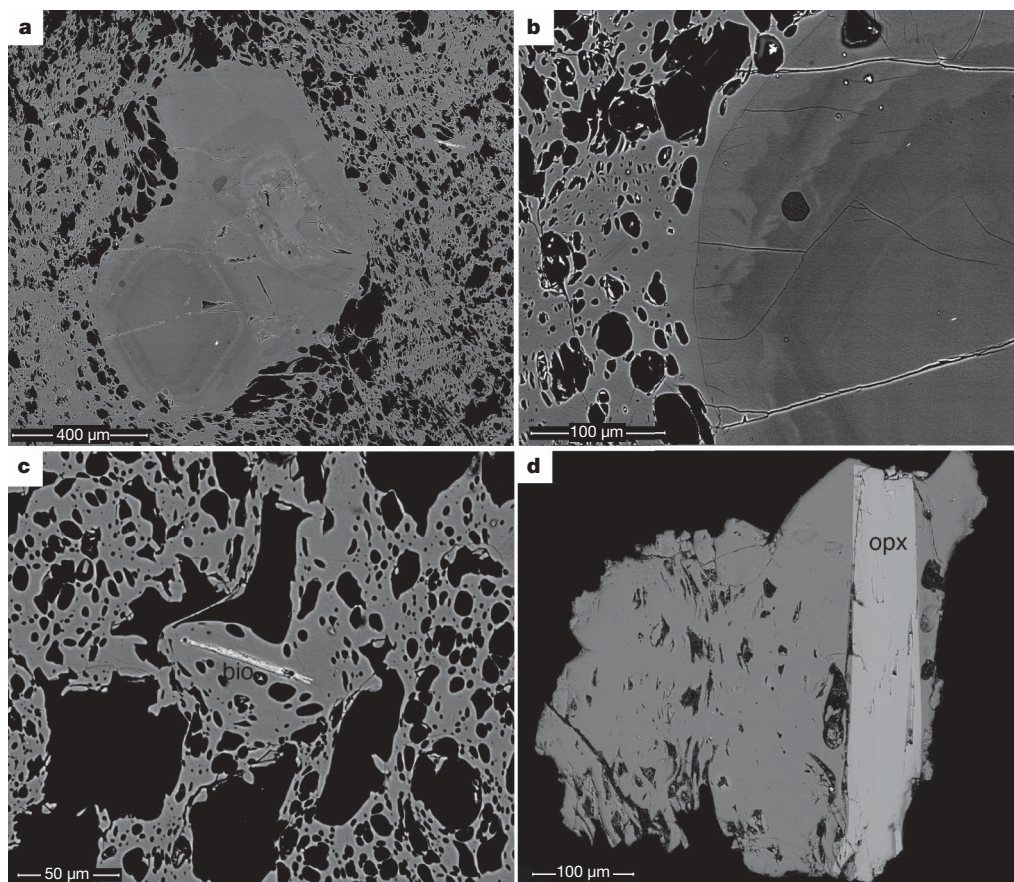


Figure 1 | Backscattered electron micrographs of Chaitén pumice pyroclasts. **a**, Plinian pumice with a plagioclase microphenocryst set within microlite-free vesicular glass. **b**, Close-up view of the lower right side of the microphenocryst in **a**. Zones of different grey-value comprise oscillatory zoning in which the anorthite content varies from about An₄₀ in dark-grey

zones to about An₄₅ in brighter regions. The dark grey blob at the centre is a hydrous rhyolite glass inclusion. **c**, Plinian pumice with biotite microlite at the grain centre. **d**, Plinian pumice pyroclast fragment with an orthopyroxene microphenocryst (light-grey elongate).

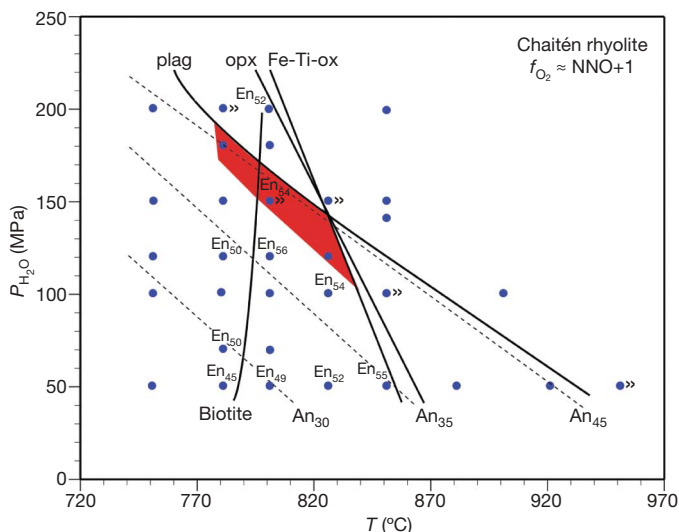


Figure 2 | H_2O -saturated phase relations in the Chaitén rhyolite. Mineral-in curves illustrate the stability limits of the natural microphenocryst minerals (plag, plagioclase; opx, orthopyroxene; Fe-Ti-ox, titanomagnetite). The symbol '»' indicates 'reversal' experiments in which pre-annealed aliquots of crystal-rich material were subjected to higher temperature. Fine dashed lines are isopleths contouring the average An-content (mol.%) in plagioclase microlites and overgrowth rims. The En-contents (mol.%) are given for selected experiments. The red region demarcates a permissible magma storage P_{H_2O} - T zone based on matching the experimental phase assemblage, mineral compositions, and crystallinity with those features observed in the Plinian pumice. f_{O_2} , oxygen fugacity.

$P > 120$ MPa. These conditions are consistent with temperatures calculated from compositions of titanomagnetite-ilmenite pairs in the pumice ($\sim 800 \pm 10^\circ C$)²¹. The pressure minimum (120 MPa) corresponds to a magma chamber depth of about 5 km (± 0.5 km),

assuming a range of country rock densities ($2,300$ – $2,700 \text{ kg m}^{-3}$). Higher magma storage pressures may be warranted, as the presence of biotite in some samples implies $T < 800^\circ C$, and consequently higher P_{H_2O} , to stabilize plagioclase of the natural composition. Taking these observations into account, a range of magma storage conditions is possible in P_{H_2O} - T space (Fig. 2).

All experiments conducted at P_{H_2O} and/or a T less than the inferred storage conditions grew abundant microlites (~ 10 – 40 vol.%), in addition to euhedral overgrowth rims on plagioclase fragments over timescales of days (Supplementary Information). The nearly aphyric character of the natural pumice is in contrast to this, and therefore must indicate magma storage at near-liquidus conditions, and then very rapid ascent to the surface. That the plagioclase microphenocrysts remained rounded during magma ascent across a P_{H_2O} - T space that should have promoted plagioclase crystallization (Fig. 2) indicates that Chaitén magma rose faster than some threshold rate that would have allowed plagioclase rim growth. To quantify this rate, we performed decompression experiments²² along temperature isotherms that bracket the range of possible storage conditions (780 and $825^\circ C$). The starting pressures (200 and 150 MPa) lie just above the plagioclase liquidus such that any microphenocryst fragments included in the annealed powder were partly resorbed before decompression. We equilibrated aliquots of the powdered pumice at the starting conditions for three days, and then decompressed charges to a final pressure of 30 MPa in a series of 5 MPa steps. The decompression intervals range from about 4 to 17 h. The dwell periods (7.5–15 min) between decompression steps define linear decompression rates of 10, 20 and 40 MPa h^{-1} .

Faceted plagioclase rims grew at 10 and 20 MPa h^{-1} at both 780 and $825^\circ C$, but did not grow at 40 MPa h^{-1} in the $825^\circ C$ experiment (Fig. 3). Very thin ($< 3 \mu\text{m}$), discontinuous rims grew in the 40 MPa h^{-1} , $780^\circ C$ run, but these rims did not grow in a subsequent experiment at 50 MPa h^{-1} . Therefore, the decompression rate that precludes the formation of plagioclase rims is similar at 780 and

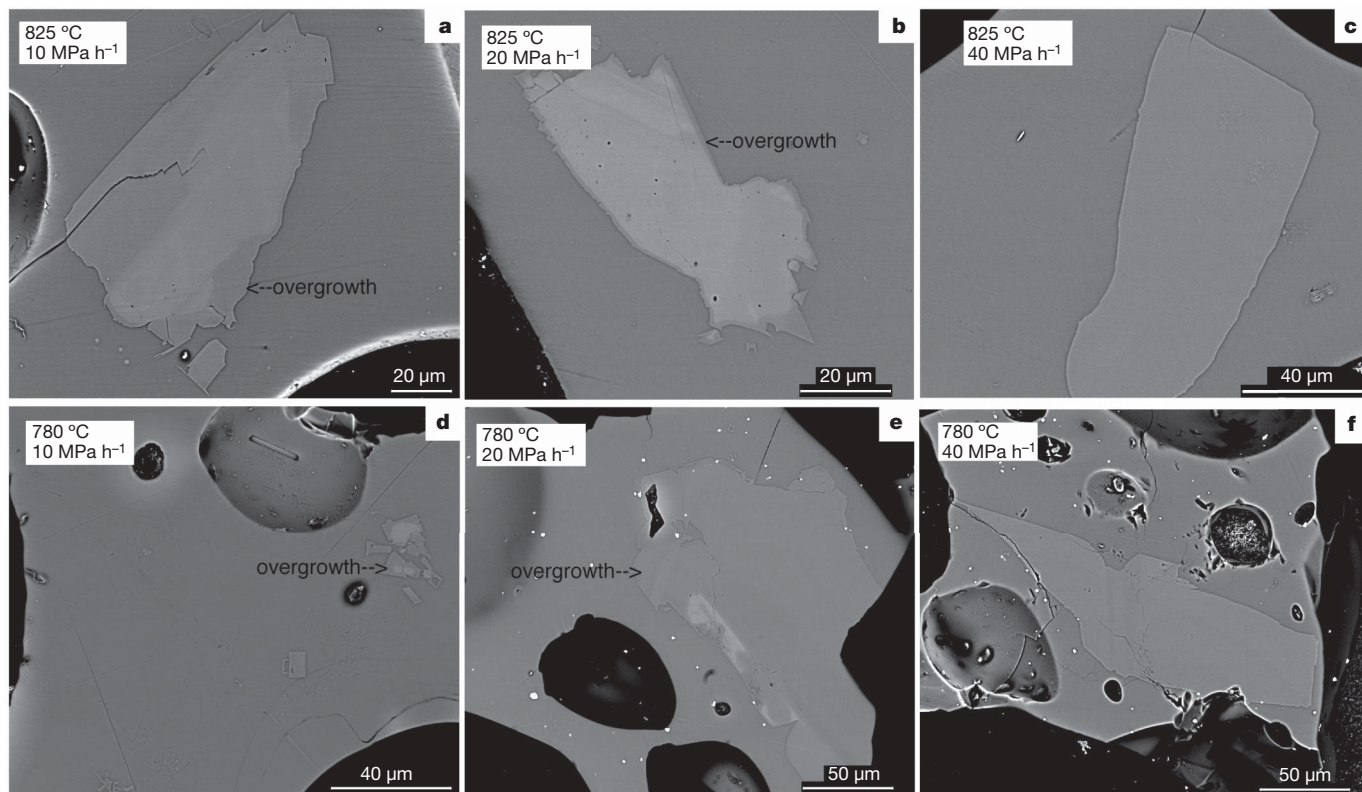


Figure 3 | Montage of backscattered electron images collected on decompression experiments on the Chaitén pumice. Each image shows one or more plagioclase microphenocrysts with an overgrowth of new plagioclase,

appearing as a darker rim (a, b, d, e) and without overgrowth (c, f). At $780^\circ C$ and $825^\circ C$, decompression rates higher than 20 MPa h^{-1} prevent the overgrowth of plagioclase owing to very short crystallization intervals.

825 °C ($\sim 40 \text{ MPa h}^{-1}$), and can be considered a minimum value for the Chaitén magma. This rate corresponds to an average ascent velocity of about 0.5 m s^{-1} , considering a bubble-free magma overburden density of about $2,300 \text{ kg m}^{-3}$.

We note that the limiting decompression rates are within the viscous regime and are therefore consistent with intact magma ascent presaging the explosive failure at higher decompression rates^{23,24}. In other words, the decompression history recorded in the pumice probably represents pre-fragmentation ascent. The ascent rate inferred from decompression experiments ($\sim 0.5 \text{ m s}^{-1}$) would correspond to strain rates ($= \text{ascent rate/conduit radius}$; $r = 10\text{--}100 \text{ m}$) of about 10^{-1} to 10^{-3} s^{-1} . The viscosity of the Chaitén magma over a range of temperature and water contents²⁵ (750–825 °C, 1–4 wt% H_2O) is still at least one order of magnitude lower ($\sim 10^6\text{--}10^8 \text{ Pa s}$) than the critical values required to cause a glassy response of the magma ($\sim 10^9\text{--}10^{10} \text{ Pa s}$) at the implied shear strain rates²⁶. This result indicates that during much of its rise in the conduit, the Chaitén magma would not have been capable of autobrecciating as a result of shear²⁷, unlike silicic magma in its final stages of ascent in lava dome eruptions^{8,9}.

Our results show for the first time that rhyolite magma can ascend very rapidly from depth ($>5 \text{ km}$) before explosive fragmentation. The magma ascent timescale at Chaitén was brief ($\sim 4 \text{ h}$), and shorter than the period of the felt seismic unrest ($\sim 1 \text{ day}$) that preceded the eruption. That the earthquake swarm duration exceeded the pre-eruptive ascent timescale may reflect preparatory fracturing and the formation of the magma's pathway to the surface²⁸, or perhaps the swarm was an eruption trigger. Nonetheless, the brevity of pre-eruptive magma rise at Chaitén is clear evidence that near-liquidus, hydrous rhyolite is very fluid and, in all likelihood, capable of creating and transiting a magma transport system²⁹ on timescales that are difficult to prepare for, especially in the absence of monitoring instruments. Our findings therefore emphasize the need to monitor rhyolite volcanoes that have undergone Holocene rhyolitic activity. In more densely populated regions this would be essential to avoid a major volcanic disaster.

METHODS SUMMARY

Glass and mineral compositions were analysed using Cameca SX-100 and JEOL JXA-8900R electron microprobes at the University of Munich and the Smithsonian Institution, respectively. Glasses were analysed with an accelerating voltage of 12–15 keV, a 10–20 μm beam, and 10 nA beam current; mineral analyses used a 3–5 μm beam and the same acceleration voltage and current. Standardization was performed on quartz (Si), anorthite (Ca), bytownite (Al), corundum (Al), microcline (K), albite (Na), hornblende (Fe, Mg), ilmenite (Ti, Fe), and chromite (Cr). Na was analysed first in all routines to minimize migration effects.

We analysed very small ($<10 \mu\text{m}$) plagioclase and orthopyroxene microlites by quantitative EDS on a FEI field emission scanning electron microscope (SEM) at the Smithsonian Institution. Analytical conditions consisted of 10–12 keV, a beam current of 0.5–1 nanoamps, spot size of $\sim 1 \mu\text{m}$, and 5 mm working distance. We calibrated the instrument against plagioclase, pyroxene, and glass standards the compositions of which were independently analysed by either electron probe microanalyser (EPMA) or wet chemistry. Reproducibility of the standard, experimental and natural mineral compositions with quantitative EDS was good, as reflected by errors of about $\pm 3.0 \text{ mol\% An}$ and $\pm 5.0 \text{ mol\% En}$ relative to the EPMA and wet chemical values.

The water contents of obsidian chips and glass inclusions were determined with synchrotron-source Fourier transform infrared (FTIR) spectroscopy at the Lawrence Berkeley National Laboratory Advanced Light Source, according to techniques described in ref. 30.

Hydrothermal phase equilibrium and decompression experiments were conducted in water-pressurized Waspaloy cold-seal vessels with nickel filler rods, according to methods described in refs 18 and 22.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 27 April; accepted 21 August 2009.

1. Wilson, C. J. N. & Walker, G. P. L. The Taupo eruption, New Zealand. I. General aspects. *Phil. Trans. R. Soc. Lond. A* **314**, 199–228 (1985).

2. Carn, S. *et al.* The unexpected awakening of Chaitén volcano, Chile. *Eos* **90**, 205–206 (2009).
3. Soosalu, H. & Einarsson, P. Earthquake activity related to the 1991 eruption of the Hekla volcano, Iceland. *Bull. Volcanol.* **63**, 536–544 (2002).
4. Kilburn, C. R. J. & Voight, B. Slow rock fracture as eruption precursor at Soufriere Hills volcano, Montserrat. *Geophys. Res. Lett.* **25**, 3665–3668 (1998).
5. Roman, D. C. & Cashman, K. V. The origin of volcano-tectonic earthquake swarms. *Geology* **34**, 457–460 (2006).
6. Chouet, B. Long-period volcano seismicity: its source and use in eruption forecasting. *Nature* **380**, 309–316 (1996).
7. Voight, B. R. & Cornelius, R. R. Prospects for eruption prediction in near real-time. *Nature* **350**, 695–698 (1991).
8. Tuffen, H., Smith, R. & Sammonds, P. Evidence for seismogenic fracture of silicic magma. *Nature* **253**, 511–514 (2008).
9. Lavallée, Y. *et al.* Seismogenic lavas and explosive eruption forecasting. *Nature* **453**, 507–510 (2008).
10. Gardner, J. E., Hilton, M. & Carroll, M. R. Experimental constraints on degassing of magma: isothermal bubble growth during continuous decompression from high pressure. *Earth Planet. Sci. Lett.* **168**, 201–218 (1999).
11. Llewellyn, E. W. & Manga, M. Bubble suspension rheology and implications for conduit flow. *J. Volcanol. Geotherm. Res.* **143**, 205–217 (2005).
12. Kilburn, C. R. J. & Sammonds, P. R. Maximum warning times for imminent volcanic eruptions. *Geophys. Res. Lett.* **32**, L24313 (2005).
13. Rutherford, M. J. in *Minerals Inclusions and Volcanic Processes* (eds Putirka, K. D. & Tepley, F. J.) 241–271 (Mineralogical Society of America, 2008).
14. Watt, S. *et al.* Fallout and distribution of volcanic ash over Argentina following the May 2008 explosive eruption of Chaitén, Chile. *J. Geophys. Res.* **114**, B04207, doi:10.1029/2008JB006219 (2009).
15. Blundy, J., Cashman, K. & Humphreys, M. Magma heating by decompression-driven crystallization beneath andesite volcanoes. *Nature* **443**, 76–80 (2006).
16. Naranjo, J. A. & Stern, C. R. Holocene tephrochronology of the southernmost part ($42^{\circ}30'\text{--}45^{\circ}\text{S}$) of the Andean Southern Volcanic Zone. *Rev. Geol. Chile* **31**, 291–306 (2004).
17. Hammer, J. E. & Rutherford, M. J. Petrologic indicators of pre-eruption magma dynamics. *Geology* **31**, 79–82 (2003).
18. Coombs, M. L. & Gardner, J. E. Shallow-storage conditions for the rhyolite of the 1912 eruption at Novarupta, Alaska. *Geology* **29**, 775–778 (2001).
19. Hammer, J. E., Rutherford, M. J. & Hildreth, W. Magma storage prior to the 1912 eruption at Novarupta, Alaska. *Contrib. Mineral. Petrol.* **144**, 144–162 (2002).
20. Hammer, J. E. & Rutherford, M. J. An experimental study of the kinetics of decompression-induced crystallization in silicic melt. *J. Geophys. Res.* **107**, doi:10.1029/2001JB000281 (2002).
21. Ghiorsio, M. S. & Evans, B. W. Thermodynamics of rhombohedral oxide solid solutions and a revision of the Fe-Ti two-oxide geothermometer and oxygen-barometer. *Am. J. Sci.* **308**, 957–1039 (2008).
22. Castro, J. M. & Gardner, J. E. Did magma ascent rate control the explosive-effusive transition at the Inyo volcanic chain, CA? *Geology* **36**, 279–282 (2008).
23. Papale, P. Strain-induced magma fragmentation in explosive eruptions. *Nature* **397**, 425–428 (1999).
24. Dingwell, D. B. Volcanic dilemma: flow or blow? *Science* **273**, 1054–1055 (1996).
25. Giordano, D., Russell, J. K. & Dingwell, D. B. Viscosity of magmatic liquids: A model. *Earth Planet. Sci. Lett.* **271**, 123–134 (2008).
26. Gonnermann, H. M. & Manga, M. Explosive volcanism may not be an inevitable consequence of magma fragmentation. *Nature* **426**, 432–435 (2003).
27. Tuffen, H., Dingwell, D. B. & Pinkerton, H. Repeated fracture and healing of silicic magma generate flow banding and earthquakes? *Geology* **31**, 1089–1092 (2003).
28. Smith, R., Kilburn, C. R. J. & Sammonds, P. R. Rock fracture as a precursor to lava dome eruptions at Mount St. Helens from June 1980 to October 1986. *Bull. Volcanol.* **69**, 681–693 (2007).
29. Scandone, R., Cashman, K. V. & Malone, S. D. Magma supply, magma ascent and the style of volcanic eruptions. *Earth Planet. Sci. Lett.* **253**, 513–529 (2007).
30. Castro, J. M. *et al.* Timescales of spherulite crystallization inferred from water concentration profiles. *Am. Mineral.* **93**, 1816–1822 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful for funding from the Alexander von Humboldt Stiftung, the Smithsonian Institution, and ERC grant 202844. N. La Penna provided an eyewitness account of the eruption sequence and critical field assistance. We thank T. Fehr, S. Bernstein and A. Logan for their analytical support. M. Rutherford and H. Tuffen provided comments that greatly improved the manuscript.

Author Contributions J.M.C. collected samples, performed the experiments and analytical work, and co-wrote the paper. D.B.D. analysed data and co-wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.M.C. (jonathan.castro@cnrs-orleans.fr).

METHODS

Pumice samples were collected on 21 June 2008 at two locations approximately 12 and 2 km from the active vent. The samples in both locations were probably erupted on 3 May 2008, based on the sample locations relative to the trajectory of wind-dispersed ash plumes as recorded by NASA satellite photos. Serial sections of the air-fall deposit were sampled across its entire thickness. Samples from all horizons were analysed to establish whether any time-variations in chemistry or mineralogy occurred in the pumice. No such variations have been detected. Samples were sieved to determine the weight fractions of the different components. Mineral identification was carried out with optical and scanning electron microscopes.

Fe-Ti oxide grains from the Plinian pumice were analysed by a Cameca SX-100 EPMA. Grains were hand-picked from the 0.5–1 mm size fraction, using magnetism and appearance of the crystals as observed under a binocular microscope. All grains were enclosed in a small selvage of vesicular glass. About 200 grains were selected in this manner, mounted in epoxy, and polished for EPMA analysis. Analytical conditions for the oxide grains included 15 keV acceleration voltage, a 20 nA beam current and focused spot. We calibrated the EPMA with well-characterized mineral standards: haematite (Fe), ilmenite (Ti, Mn), chromite (Cr), corundum (Al) and periclase (Mg). Analyses of unknowns were interspersed with measurements of the mineral standards to check for instrument drift.

Fe-Ti oxide compositional data was used to calculate the temperature and oxygen fugacity of the Chaitén rhyolite. Of the 200 Fe-Ti oxide grains analysed, only five titanomagnetite-ilmenite pairs were found to be in contact with one another, and hence suitable for estimating intensive parameters. We tested that these oxide pairs had been in equilibrium with each other by analysing Mg/Mn and using the empirical model of ref. 31. All pairs were found to be in equilibrium based on their Mg/Mn values (Supplementary Information Table 4). We used the two-oxide solid solution model of ref. 21 to calculate temperature and f_{O_2} .

Glass, plagioclase and pyroxene compositions were analysed using Cameca SX-100 and JEOL JXA-8900R electron microprobes at the University of Munich and the Smithsonian Institution, respectively. Glasses were analysed with an accelerating voltage of 12–15 keV, a 10–20 μm beam and 10 nA beam current; mineral analyses used a 3–5 μm beam and the same acceleration voltage and current. Standardization was performed on quartz (Si), anorthite (Ca), bytownite (Al), microcline (K), albite (Na), hornblende (Fe, Mg), ilmenite (Ti, Fe), and chromite (Cr). Na was analysed first in all routines to minimize migration effects.

We analysed some very small (<10 μm) plagioclase and orthopyroxene microlites by quantitative energy dispersive spectroscopy (EDS) on a FEI field emission SEM at the Smithsonian Institution. Analytical conditions comprised: 10–12 keV, a beam current of 0.5–1 nanoamps, spot size of $\sim 1 \mu\text{m}$, and 5 mm working distance. We calibrated the instrument against plagioclase, pyroxene, and glass standards the compositions of which were independently analysed by either EPMA or wet chemistry. Reproducibility of the standard, experimental, and natural mineral compositions with quantitative EDS was good, as reflected by errors of about +3.0 mol% An and +5.0 mol% En relative to the EPMA and wet chemical values.

The water contents of obsidian chips and glass inclusions were determined with synchrotron-source FTIR at the Lawrence Berkeley National Laboratory Advanced Light Source, according to techniques described in ref. 29. All measurements were made in transmission mode on doubly polished wafers ranging in thickness from 30 to 200 μm .

Hydrothermal experiments on a powdered aliquot of Chaitén pumice were conducted in water-pressurized Waspalloy cold-seal vessels with nickel filler rods, according to methods described in refs 16 and 21. All experiments were run at an ambient oxygen fugacity of approximately NNO+1 log unit. Pumice powder, along with enough distilled water to ensure H_2O -saturation ($P_{\text{H}_2\text{O}} = P_{\text{total}}$) at the run conditions were loaded into Au tubes and welded shut. Capsules were weighed before and after the experiments to confirm that they remained closed during the runs. Experiments that underwent weight loss were discarded. Experiments were quenched first with air and then by immersion in a cold-water bath.

Decompression experiments were first equilibrated at the starting conditions (200 MPa, 750 °C; 150 MPa, 825 °C) for 1 to 3 days. Runs were then decompressed by a series of equal-sized steps (5 MPa) with a hand-operated pressure intensifier. The size of the intervening dwell periods at the intermediate pressures established three model decompression rates (10, 20 and 40 MPa h^{-1}). The final quench pressure in decompression experiments (30 MPa) was chosen as an intermediate to the pressures implied by the water contents of obsidian pyroclasts and the plagioclase-hosted glass inclusions, assuming that these water contents reflect equilibrium solubility values³².

31. Bacon, C. R. & Hirschmann, M. M. Mg/Mn partitioning as a test for equilibrium between coexisting Fe-Ti oxides. *Am. Mineral.* **73**, 57–61 (1988).
32. Silver, L. A., Ihinger, P. D. & Stolper, E. The influence of bulk composition on the speciation of water in silicate glasses. *Contrib. Mineral. Petrol.* **104**, 142–162 (1989).

LETTERS

Gene therapy for red–green colour blindness in adult primates

Katherine Mancuso¹, William W. Hauswirth², Qihong Li², Thomas B. Connor³, James A. Kuchenbecker¹, Matthew C. Mauck³, Jay Neitz¹ & Maureen Neitz¹

Red–green colour blindness, which results from the absence of either the long- (L) or the middle- (M) wavelength-sensitive visual photopigments, is the most common single locus genetic disorder. Here we explore the possibility of curing colour blindness using gene therapy in experiments on adult monkeys that had been colour blind since birth. A third type of cone pigment was added to dichromatic retinas, providing the receptor basis for trichromatic colour vision. This opened a new avenue to explore the requirements for establishing the neural circuits for a new dimension of colour sensation. Classic visual deprivation experiments¹ have led to the expectation that neural connections established during development would not appropriately process an input that was not present from birth. Therefore, it was believed that the treatment of congenital vision disorders would be ineffective unless administered to the very young. However, here we show that the addition of a third opsin in adult red–green colour-deficient primates was sufficient to produce trichromatic colour vision behaviour. Thus, trichromacy can arise from a single addition of a third cone class and it does not require an early developmental process. This provides a positive outlook for the potential of gene therapy to cure adult vision disorders.

Gene therapy was performed on adult squirrel monkeys (*Saimiri sciureus*) that were missing the L-opsin gene. In this species, some females have trichromatic colour vision whereas males are red–green colour blind². Serotype 2/5 recombinant adeno-associated virus (rAAV) containing a human L-opsin gene under the control of the L/M-opsin enhancer and promoter (Fig. 1a) was delivered to the photoreceptor layer by subretinal injections (see Methods). Transcriptional regulatory elements were chosen to direct expression preferentially in M cones, but not short- (S) wavelength-sensitive cones or rods³. To provide the receptor basis for trichromacy, animals received three 100- μ l injections (containing a total of 2.7×10^{13} viral particles) in each eye, which produced a relatively uniform, third submosaic of approximately 15–36% of M cones that coexpressed the transgene (Fig. 1e, f).

Before treatment, monkeys were trained to perform a computer-based colour vision test, the Cambridge Colour Test^{4,5}, which was modified for use with animals⁶ (Fig. 2a). Dichromats who are missing either the L- or the M-photopigment fail to distinguish from grey: colours near the so-called ‘spectral neutral point’ located in the blue–green region of colour space (near dominant wavelength of 490 nm) and complementary colours near the ‘extra-spectral neutral point’ in the red–violet region (near dominant wavelength of \sim 499 nm). Whereas trichromats have the four main hue percepts blue, yellow, red and green, dichromats only have two percepts, nominally blue and yellow. Before treatment, two dichromatic monkeys completed

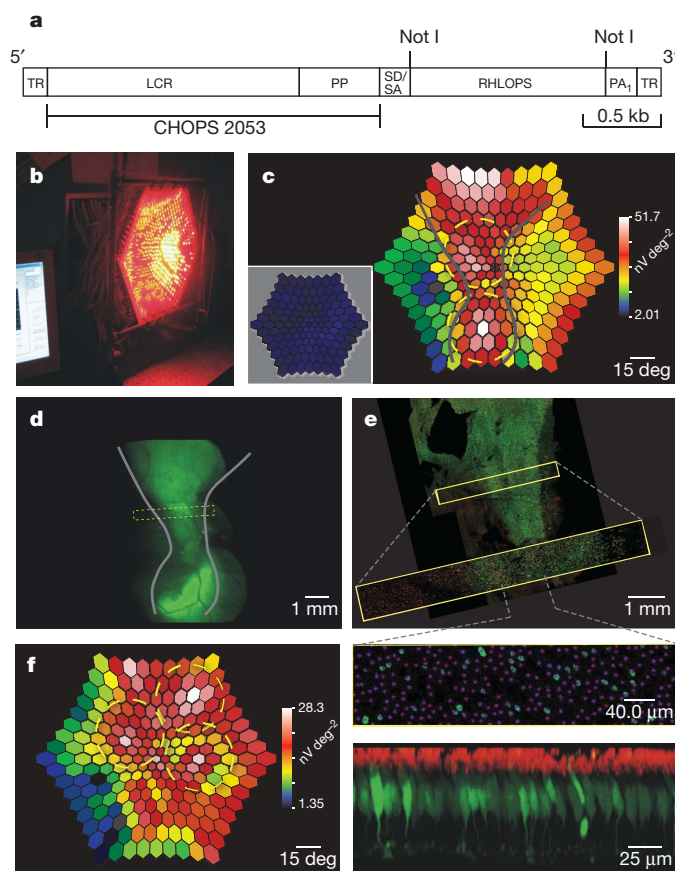


Figure 1 | rAAV2/5 vector produced functional L-opsin in primate retina.

a, Molecular map. LCR, locus control region; PA₁, polyadenylation signal; PP, proximal promoter; RHLOPS, recombinant human L-opsin cDNA; SD/SA, splice donor/acceptor; TR, terminal repeats. **b**, Red light mf-ERG stimulus. **c**, mf-ERG 40 weeks after two injections (yellow circles) of a mixture of L-opsin- and GFP-coding viruses. Grey lines show borders of highest response. For comparison, the inset shows mf-ERG 16 weeks after injection; there was no reliable signal from L-opsin, unchanged from baseline. High responses in far peripheral retina were measured reliably and may have originated from offshoot of one of the injections. **d**, Fluorescence photographs from a similar retinal area as in **c**; grey lines from **c** were copied in **d**. **e**, Confocal microscopy showed a mosaic pattern of GFP expression in 5–12% of cones. Because GFP-coding virus was diluted to one-third compared to L-opsin virus, an estimated 15–36% of cones in behaviourally tested animals express L-opsin. **f**, Mf-ERG from a behaviourally tested animal 70 weeks after three injections of L-opsin virus.

¹Department of Ophthalmology, Box 356485, University of Washington, 1959 North East Pacific Street, Seattle, Washington 98195, USA. ²Department of Ophthalmology and Powell Gene Therapy Center, University of Florida, 1600 South West Archer Road, Gainesville, Florida 32610, USA. ³Department of Ophthalmology, Medical College of Wisconsin, 925 North 87th Street, Milwaukee, Wisconsin 53226, USA.

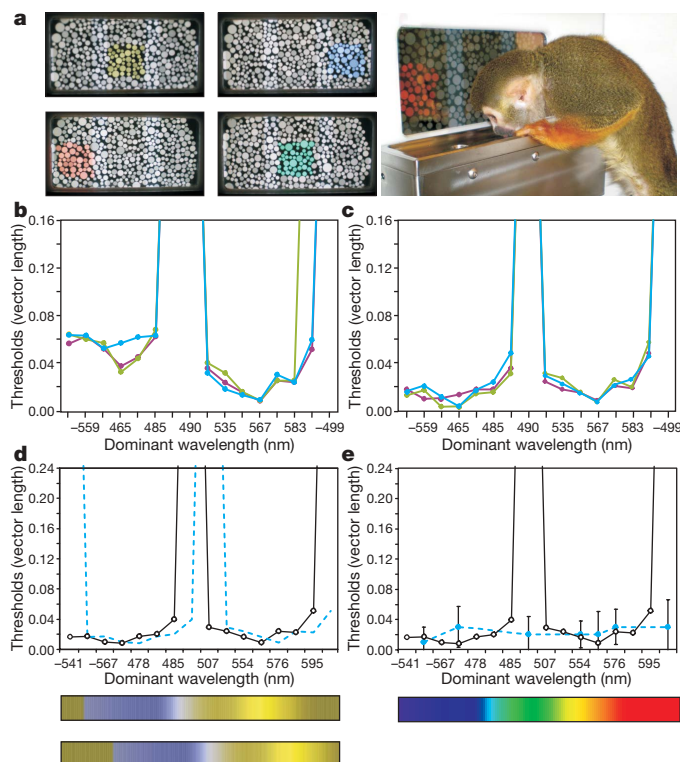


Figure 2 | Pre-therapy colour vision and possible treatment outcomes.

a, Colour-vision stimuli examples. **b**, Pre-therapy results, monkey 1. Hues tested are represented as dominant wavelengths rather than u' , v' coordinates. If a hue could not be reliably distinguished at even the highest saturation, the extrapolated threshold approached infinity. **c**, Pre-therapy results, monkey 2. **d**, **e**, Possible experimental outcomes: monkeys could have a relative increase in long-wavelength sensitivity, but remain dichromatic (dashed lines, **d**); theoretical colour spectrum appearances for a dichromat and a possible 'spectral shift' are shown. Alternatively, dichromatic monkeys could become trichromatic. Results from a trichromatic female control monkey are plotted (dashed line, **e**). Error bars denote s.e.m.; n varied from 7–11.

three colour vision tests consisting of 16 hues (Fig. 2b, c). Four-to-six months were required to test all 16 hues; thus, baseline results represent testing conducted for more than a year. As predicted, before treatment monkeys had low thresholds (averaging <0.03 units in u' , v' colour space) for colours that represent blues and yellows to their eyes, but always failed to discriminate between the blue-green and the red-violet (dominant wavelengths of 490 nm and -499 nm, respectively) hues, with thresholds extrapolated from psychometric functions being orders of magnitude higher (Fig. 2b, c). Results were highly repeatable, with no improvement between the first and third tests, making us confident that the animals would not spontaneously improve in the absence of treatment.

Co-expressing the L-opsin transgene within a subset of endogenous M-cones shifted their spectral sensitivity to respond to long wavelength light, thus producing two distinct cone types absorbing in the middle-to-long wavelengths, as required for trichromacy. The spectral sensitivity shift was readily detected using a custom-built wide-field colour multifocal electroretinogram (mf-ERG) system (Fig. 1b, c, f) (see ref. 7 for details). In preliminary experiments, validity of the colour mf-ERG was tested using an animal that had received a mixture of the L-opsin-coding virus plus an identical virus, except that a green fluorescent protein (GFP) gene replaced the L-opsin gene. As reported previously, faint GFP fluorescence was first detected at 9 weeks post-injection, and it continued to increase in area and intensity over 24 weeks⁸. Although faint signs of GFP were first detectable at 9 weeks, L-opsin levels sufficient to produce

suprathreshold mf-ERG signals were still not present at 16 weeks post-injection (Fig. 1c, inset). After GFP fluorescence became robust, the red light mf-ERG, which indicates responses from the introduced L-opsin, showed highly increased response amplitudes in two areas (Fig. 1c) corresponding to locations of subretinal injections (Fig. 1d).

The two dichromatic monkeys who participated in behavioural tests of colour vision were treated with L-opsin-coding virus only. Although the elongated pattern produced by two injections in Fig. 1c, d allowed mf-ERG validation, the treatment goal was to produce a homogeneous region, as resulted from three injections shown in Fig. 1f, in which the highest mf-ERG response covered about 80° of the central retina—roughly the area for which humans have good red–green discrimination. These results demonstrate that gene therapy changed the spectral sensitivity of a subset of the cones. A priori, there were two possibilities for how a change in spectral sensitivity might change colour vision behaviour. First, animals may have an increase in sensitivity to long-wavelength light, but if the neural circuitry for extracting colour information from the nascent 'M + L cone' submosaic was absent, they would remain dichromatic—the hallmark of which is having two hues that are indistinguishable from grey (Fig. 2d). The spectral neutral point for individuals that have only S and M cones (for example, monkeys 1 and 2 pre-therapy) occurs near the dominant wavelength of 495 nm. At the limit, an increase in spectral sensitivity would shift the monkeys' neutral point towards that of individuals with only S and L cones, near the dominant wavelength of 505 nm (Fig. 2d, dashed blue lines). The second, more engaging possibility was that treatment would be sufficient to expand sensory capacity in monkeys, providing them with trichromatic vision. In this case, the animals' post-therapy results would appear similar to Fig. 2e, obtained from a trichromatic female control monkey.

Daily testing continued after treatment. After about 20 weeks post-injection (Fig. 3a, arrow), the trained monkeys' thresholds for blue-green and red-violet (dominant wavelengths of 490 and -499 nm, respectively; Fig. 3b, c) improved, reducing to an average of 0.08 units in u' , v' colour space, indicating that they gained trichromatic vision. This time point corresponded to the same period in which robust levels of transgene expression were reported in the squirrel monkey⁸. A trichromatic female monkey and untreated dichromatic monkeys were tested in parallel. As expected, the female had low thresholds for all colours, averaging <0.03 units in u' , v' colour space, but the untreated dichromats always failed to discriminate between dominant wavelengths of 490 nm (Fig. 3a, triangle) and -499 nm, indicating a clear difference between treated and untreated monkeys.

Early experiments in which we obtained negative results served as 'sham controls', demonstrating that acquiring a new dimension of colour vision requires a shift in spectral sensitivity that results from expression of an L pigment in a subset of M cones. Using similar subretinal injection procedures, we delivered fewer viral particles of an L-opsin-coding rAAV2/5 virus with an extra 146-base-pair (bp) segment near the splice donor/acceptor site that had been carried over from the cloning vector and that was absent in the GFP-coding rAAV2/5 virus. The 146-bp segment contained an ATG and a duplicate messenger RNA start site that may have interfered with expression (see Methods). Three monkeys received injections of this vector, containing an average of 1.7×10^{12} virus particles per eye, and no reliable changes in spectral sensitivity were measured using the ERG. One animal was also tested behaviourally and his colour vision was unchanged from baseline 1 year after injection. In subsequent experiments reported here, we removed the extra 146-bp segment and also increased the amount of viral particles delivered per eye by approximately 16-fold, to 2.7×10^{13} . Negative results from earlier injections demonstrated that the subretinal injection procedure itself does not produce changes in the ERG or in colour vision.

The change in spectral sensitivity measured with the mf-ERG is necessary but not sufficient to produce a new colour vision capacity. For example, individuals with L but no M cones (termed deuteranopes) have a relatively enhanced sensitivity to red light, but they are still as

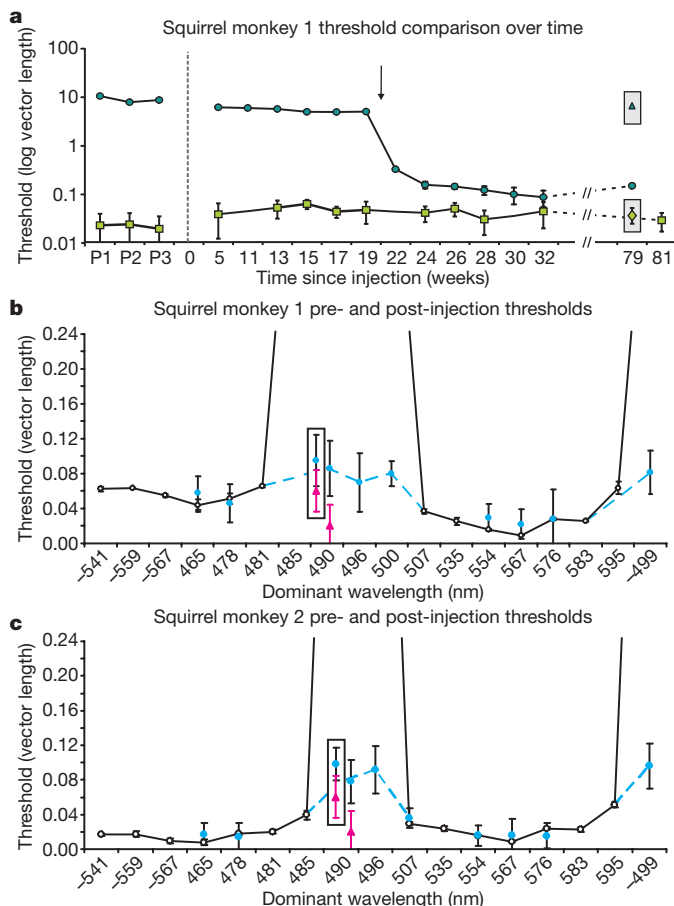


Figure 3 | Gene therapy produced trichromatic colour vision. **a**, Time course of thresholds for the blue-green confusion colour, dominant wavelength of 490 nm (circles), and a yellowish colour, dominant wavelength of 554 nm (squares). A logarithmic scale was used to fit high thresholds for the dominant wavelength of 490 nm; significant improvement occurred after 20 weeks. Enclosed data points denote untreated dichromatic monkey thresholds, dominant wavelengths of 490 nm (triangle) and 554 nm (diamond). **b**, **c**, Comparison of pre-therapy (open circles, solid line) and post-therapy (solid dots, dashed line) thresholds. Enclosed data points are dominant wavelength 490 nm thresholds when tested against a red-violet background (dominant wavelength of -499 nm); pink triangles show trichromatic female control thresholds. Error bars represent s.e.m.; n varied from 7–11.

dichromatic as individuals with M but no L cones (protanopes) in that they are unable to distinguish particular 'colours' from grey. To verify that the behavioural change observed in animals expressing the L pigment transgene was not purely a shift in spectral sensitivity (see Fig. 2d), monkey 1 was also tested on dominant wavelengths of 496 and 500 nm, and monkey 2 was tested on dominant wavelengths of 496 and 507 nm. Together, these dominant wavelengths span the possible confusion points for deuteranopes and protanopes and for any intermediate dichromatic forms that could arise from expressing combinations of L and M pigments. As shown in Fig. 3b, c, both monkeys' measured thresholds for these extra hues were similar to their thresholds for a dominant wavelength of 490 nm, demonstrating that they now lacked a spectral neutral point and have become truly trichromatic. Furthermore, treated monkeys were able to discriminate blue-green (dominant wavelength of 490 nm) when it was tested against a red-violet (dominant wavelength of -499 nm) background, instead of the grey background, indicating that the monkeys' newly-acquired 'green' and 'red' percepts were distinct from one another. The treated monkeys' improvement in colour vision has remained stable for more than 2 years and we plan to continue testing the animals to evaluate long-term treatment effects.

Classic experiments in which visual deprivation of one eye during development caused permanent vision loss¹ led to the idea that inputs must be present during development for the formation of circuits to process them. From the clear change in behaviour associated with treatment, compared both between and within subjects, we conclude that adult monkeys gained new colour vision capacities because of gene therapy. These startling empirical results provide insight into the evolutionary question of what changes in the visual system are required for adding a new dimension of colour vision. Previously, it seemed possible that a transformation from dichromacy to trichromacy would require evolutionary/developmental changes, in addition to acquiring a third cone type. For example, L- and M-opsin-specific genetic regulatory elements might have been required to direct the opsins into distinct cone types⁹ that would be recognized by L- and M-cone-specific retinal circuitry¹⁰, and to account for cortical processing, multi-stage circuitry¹¹ might have evolved specifically for the purpose of trichromacy. However, our results demonstrate that trichromatic colour vision behaviour requires nothing more than a third cone type. As an alternative to the idea that the new dimension of colour vision arose by acquisition of a new L versus M pathway, it is possible that it exploited the pre-existing blue-yellow circuitry. For example, if the addition of the third cone class split the formerly S versus M receptive fields into two types with differing spectral sensitivities, this would obviate the need for neural rewiring as part of the process of adopting new colour vision.

Some form of inherent plasticity in the mammalian visual system can be inferred from the acquisition of new colour vision, as was also demonstrated in genetically engineered mice¹²; however, the point has been made that such plasticity need not indicate that any rewiring of the neural circuitry has occurred¹³. Similarly, given the fact that new colour vision behaviour in adult squirrel monkeys corresponded to the same time interval as the appearance of robust levels of transgene expression, we conclude that rewiring of the visual system was not associated with the change from dichromatic to trichromatic vision.

Treated adult monkeys unquestionably respond to colours that were previously invisible to them. The internal experiences associated with the marked change in discrimination thresholds measured here cannot be determined; therefore, we cannot know whether the animals experience new internal sensations of red and green. Nonetheless, we do know that evolution acts on behaviour, not on internalized experiences, and we suggest that gene therapy recapitulated what occurred during evolution of trichromacy in primates. These experiments demonstrate that a new colour-vision capacity, as defined by new discrimination abilities, can be added by taking advantage of pre-existing neural circuitry and, internal experience aside, full colour vision could have evolved in the absence of any other change in the visual system except the addition of a third cone type.

Gene therapy trials are underway for Leber's congenital amaurosis^{14–16}. Thus far, treatment has been administered to individuals who have suffered retinal degeneration from the disease. The experiments reported here are, to our knowledge, the first to use gene therapy in primates to address a vision disorder in which all photoreceptors are intact and healthy, making it possible to assess the full potential of gene therapy to restore visual capacities. Treatment allowing monkeys to see new colours in adulthood provides a striking counter-example to what occurs under conditions of monocular deprivation. For instance, it is impossible to restore vision in an adult who had grown up with a unilateral cataract. Future technologies will allow many opportunities for functions to be added or restored in the eye. Although some changes may produce outcomes analogous to monocular deprivation, we predict that others, like gene therapy for red-green colour blindness, will provide vision where there was previously blindness.

METHODS SUMMARY

Confocal microscopy. The animal in Fig. 1c, d succumbed to respiratory illness, unrelated to gene therapy, approximately 2 years and 3 months after injection.

The retina was fixed in 4% paraformaldehyde in PBS, and rinsed in PBS with 10% and 30% sucrose. It was sequentially incubated with 10% normal donkey serum, rabbit monoclonal antibody to M/L-opsin (Chemicon, AB5405), and a Cy3 (red)-conjugated donkey anti-rabbit antibody (Jackson ImmunoResearch). Confocal images were analysed using ImageJ (<http://rsbweb.nih.gov>). In the middle panel of Fig. 1e, magenta dots mark cone locations, and the red anti-M/L-opsin antibody staining was removed to show GFP-expressing (green) cells more clearly.

Behavioural colour vision assessment. A three-alternative forced-choice model in which position and saturation of the stimulus was randomized between trials was used. Monkeys had to discriminate the location of a coloured patch of dots that varied in size and brightness, surrounded by similarly varying grey dots. When animals touched the coloured target, a positive tone sounded and a juice reward was given; the next stimulus appeared immediately. (The squirrel monkey shown in Fig. 2c is drinking a reward from a previous trial). If the wrong position was chosen, a negative tone sounded, and a 2–3-s ‘penalty time’ occurred before the next trial.

For each hue, monkeys were tested on up to 11 different saturations ranging from 0.01 to 0.11 in u' , v' colour space (CIE 1976) and a threshold was calculated, which was taken as the saturation required to reach a criterion of 57% correct, the value determined to be significantly greater than chance (33% correct, $P = 0.05$); see ref. 6 for full details. All procedures were conducted in accordance with the guidelines of the US National Institutes of Health about the care and use of animals.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 19 June; accepted 14 August 2009.

Published online 16 September 2009.

- Wiesel, T. N. & Hubel, D. H. Single-cell responses in striate cortex of kittens deprived of vision in one eye. *J. Neurophysiol.* **26**, 1003–1017 (1963).
- Jacobs, G. H. A perspective on color vision in platyrrhine monkeys. *Vision Res.* **38**, 3307–3313 (1998).
- Li, Q., Timmers, A. M., Guy, J., Pang, J. & Hauswirth, W. W. Cone-specific expression using a human red opsin promoter in recombinant AAV. *Vision Res.* **48**, 332–338 (2007).
- Reffin, J. P., Astell, S. & Mollon, J. D. in *Colour Vision Deficiencies X* (eds Drum, B., Moreland, J. D. and Serra, A.) 69–76 (Kluwer Academic Publishers, 1991).
- Regan, B. C., Reffin, J. P. & Mollon, J. D. Luminance noise and the rapid determination of discrimination ellipses in colour deficiency. *Vision Res.* **34**, 1279–1299 (1994).
- Mancuso, K., Neitz, M. & Neitz, J. An adaptation of the Cambridge Colour Test for use with animals. *Vis. Neurosci.* **23**, 695–701 (2006).
- Kuchenbecker, J. A., Sahay, M., Tait, D. M., Neitz, M. & Neitz, J. Topography of the long- to middle-wavelength sensitive cone ratio in the human retina assessed with a wide-field color multifocal electroretinogram. *Vis. Neurosci.* **25**, 301–306 (2008).
- Mancuso, K. *et al.* Recombinant adeno-associated virus targets passenger gene expression to cones in primate retina. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* **24**, 1411–1416 (2007).
- Nathans, J., Piantanida, T. P., Eddy, R. L., Shows, T. B. & Hogness, D. S. Molecular genetics of inherited variation in human color vision. *Science* **232**, 203–210 (1986).
- Shapley, R. Specificity of cone connections in the retina and color vision. Focus on ‘‘Specificity of cone inputs to macaque retinal ganglion cells’’. *J. Neurophysiol.* **95**, 587–588 (2006).
- De Valois, R. L. & De Valois, K. K. A multi-stage color model. *Vision Res.* **33**, 1053–1065 (1993).
- Jacobs, G. H., Williams, G. A., Cahill, H. & Nathans, J. Emergence of novel color vision in mice engineered to express a human cone photopigment. *Science* **315**, 1723–1725 (2007).
- Makous, W. Comment on ‘‘emergence of novel color vision in mice engineered to express a human cone photopigment’’. *Science* **318**, 196 (2007).
- Maguire, A. M. *et al.* Safety and efficacy of gene transfer for Leber’s congenital amaurosis. *N. Engl. J. Med.* **358**, 2240–2248 (2008).
- Bainbridge, J. W. & Ali, R. R. Success in sight: the eyes have it! Ocular gene therapy trials for LCA look promising. *Gene Ther.* **15**, 1191–1192 (2008).
- Cideciyan, A. V. *et al.* Human gene therapy for RPE65 isomerase deficiency activates the retinoid cycle of vision but with slow rod kinetics. *Proc. Natl Acad. Sci. USA* **105**, 15112–15117 (2008).

Acknowledgements This work was supported by the National Institutes of Health grants R01EY016861 (M.N.) and R01EY11123 (W.W.H.); Research Training Program in Vision Science Grant T32EY014537; NEI Core Grants for Vision Research P30EY01931, P30EY01730 and P30EY08571; the Harry J. Heeb Foundation, the Posner Foundation, the Macular Vision Research Foundation, the Foundation Fighting Blindness, Hope for Vision, and Research to Prevent Blindness. We would like to thank V. Chiodo, S. Boye, D. Conklyn, P. M. Summerfelt, K. Chmielewski and K. L. Gunther for technical assistance. J.N. is the Bishop Professor in Ophthalmology, M.N. is the Ray Hill Professor in Ophthalmology, and W.W.H. is Rybackzi-Bullard Professor of Ophthalmology.

Author Contributions Experiments and data analysis were performed by K.M., T.B.C., J.A.K., M.C.M., J.N. and M.N. Cone-specific expression of the gene therapy vector was developed and validated by Q.L., and W.W.H. constructed the vector and packaged it into adeno-associated virus and provided dosage guidance. All authors contributed to data interpretation. The manuscript was written by K.M., J.N. and M.N. and incorporates comments by all others.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.N. (jneitz@uw.edu).

METHODS

Viral vector. CHOPS2053 was a 2.1-kilobase (kb) fragment containing the locus control region and proximal promoter upstream of the human X-chromosome opsin gene array^{9,17}. These elements (also known as pR2.1) have been shown to target transgene expression to mammalian L/M cones^{3,18}. RHLOPS was a 1.2-kb fragment containing recombinant human L-opsin cDNA. A clone of the human L-opsin cDNA¹⁹, known as hs7, was generously provided by J. Nathans. The QuickChange kit (Stratagene) was used to convert codon 180 so that it would encode a human L pigment maximally sensitive to 562 nm²⁰. The virus was made using the genome from rAAV serotype 2 and the capsid from serotype 5, and the preparation had 9×10^{13} DNase-resistant vector genome containing particles per ml. To prevent vector aggregation, 0.014% Tween-20 was added to the final vector preparation. A total of 2.7×10^{13} viral particles were injected per eye.

An earlier version of the L-opsin-coding rAAV2/5 used in previous unsuccessful experiments contained an extra 146-bp segment between the splice donor/acceptor site and the translational start codon of the L-opsin gene that had been carried over from the cloning vector. Because we were concerned that this fragment may have interfered with transgene expression, a second version of L-opsin rAAV2/5 in which the extra 146 bp had been removed was used in later experiments described here. In addition to modifying the vector, we also increased the amount of viral particles delivered per eye by approximately 16-fold, from 1.7×10^{12} to 2.7×10^{13} . Thus, we cannot conclude from this set of experiments what exact titre of viral particles was required to produce the effects on colour vision behaviour, or exactly what effects, if any, the extra 146 bp had on transgene expression in earlier unsuccessful attempts.

The single-stranded DNA genome of conventional rAAV vectors, including rAAV2/5 used here, is devoid of *Rep* coding sequences. Thus, the vector genome is stabilized predominantly in an episomal form; however, the potential for integration exists²¹. According to NIH guidelines, the viral vector used here is rated biosafety level 1 (BSL1), and animal biosafety level 1 (ABSL1) meaning that no special precautions were required in handling the virus or animals treated with the virus. After treatment, squirrel monkeys had an increase in AAV antibody titres, ranging from 4–12-fold. Antibody titres remained unchanged in untreated control animals who were housed with treated animals.

Subretinal injections. Subretinal injections were performed by a vitreo-retinal surgeon (T.B.C.) using a KDS model 210 syringe pump under a stereomicroscope. A 500- μ l Hamilton Gastight (1750TTL) Luer Lock syringe was connected

to 88.9 cm of 30 gauge teflon tubing with male Luer Lock adapters at both ends (Hamilton 30TF double hub), which was then connected to a 30-gauge Becton Dickinson Yale regular bevel cannula (ref 511258) that was manually bent to produce a 135° angle 1.5-mm from the tip. All components were sterilized before use. The syringe and tubing were filled with sterile lactated Ringers solution to produce a dead volume of approximately 210 μ l. Just before injection, 300 μ l of rAAV was withdrawn using a rate of 100 μ l min⁻¹.

Squirrel monkeys were anaesthetized using intramuscular injections of ketamine (15 mg kg⁻¹) and xylazine (2 mg kg⁻¹); atropine (0.05 mg kg⁻¹) was also given to reduce airway secretions. The eye was dilated with 2–3 drops of tropicamide (1%) and treated with one drop each of betadine (5%), vigamox (0.5%) and proparacaine (1%). Subconjunctival injection of 0.1 ml lidocaine (2%) was given, and the anterior portion of the eye was exposed by performing a temporal canthotomy followed by limited conjunctival peritomy. Eyelids were held open with a speculum designed for premature infants. A temporal sclerotomy was made 1-mm posterior to the limbus with a 27-gauge needle, through which the injection cannula was inserted. Three subsequent 100- μ l injections were made at different subretinal locations using an infusion rate of 1,060 μ l min⁻¹. Post-procedure, 0.05 ml each of decadron (10 mg ml⁻¹), kenalog (40 mg ml⁻¹) and cephalosin (100 mg ml⁻¹) were injected subconjunctivally; one drop each of betadine (5%) and vigamox (0.5%) and a 0.6-cm strip of tobradex (0.3% tobramycin, 0.1% dexamethasone) ointment were applied topically; 10–20 ml of subcutaneous fluids (sterile lactated Ringers) was also given. Subsequent administration of steroids and analgesics was administered as needed post-procedure for potential inflammation or discomfort.

17. Wang, Y. *et al.* A locus control region adjacent to the human red and green visual pigment genes. *Neuron* **9**, 429–440 (1992).
18. Mauck, M. C. *et al.* Longitudinal evaluation of expression of virally delivered transgenes in gerbil cone photoreceptors. *Vis. Neurosci.* **25**, 273–282 (2008).
19. Nathans, J., Thomas, D. & Hogness, D. S. Molecular genetics of human color vision: the genes encoding blue, green, and red pigments. *Science* **232**, 193–202 (1986).
20. Neitz, M., Neitz, J. & Jacobs, G. H. Spectral tuning of pigments underlying red-green color vision. *Science* **252**, 971–974 (1991).
21. Büning, H., Perabo, L., Coutelle, O., Quadts-Humme, S. & Hallek, M. Recent developments in adeno-associated virus vector technology. *J. Gene Med.* **10**, 717–733 (2008).

LETTERS

STING regulates intracellular DNA-mediated, type I interferon-dependent innate immunity

Hiroki Ishikawa¹, Zhe Ma¹ & Glen N. Barber¹

The innate immune system is critical for the early detection of invading pathogens and for initiating cellular host defence countermeasures, which include the production of type I interferon (IFN)^{1–3}. However, little is known about how the innate immune system is galvanized to respond to DNA-based microbes. Here we show that STING (stimulator of interferon genes) is critical for the induction of IFN by non-CpG intracellular DNA species produced by various DNA pathogens after infection⁴. Murine embryonic fibroblasts, as well as antigen presenting cells such as macrophages and dendritic cells (exposed to intracellular B-form DNA, the DNA virus herpes simplex virus 1 (HSV-1) or bacteria *Listeria monocytogenes*), were found to require STING to initiate effective IFN production. Accordingly, *Sting*-knockout mice were susceptible to lethal infection after exposure to HSV-1. The importance of STING in facilitating DNA-mediated innate immune responses was further evident because cytotoxic T-cell responses induced by plasmid DNA vaccination were reduced in *Sting*-deficient animals. In the presence of intracellular DNA, STING relocated with TANK-binding kinase 1 (TBK1) from the endoplasmic reticulum to perinuclear vesicles containing the exocyst component Sec5 (also known as EXOC2). Collectively, our studies indicate that STING is essential for host defence against DNA pathogens such as HSV-1 and facilitates the adjuvant activity of DNA-based vaccines.

Nucleic acid species inadvertently generated by microbes after infection are potent inducers of cellular innate immune defences important for protection of the host^{1–3}. Although considerable progress has been made into unravelling how RNA viruses induce type I IFN, required for triggering the production of anti-viral genes, little is known at the molecular level about the induction of IFN by DNA pathogens such as herpes simplex virus 1 (HSV-1) or by intracellular bacteria or parasites^{5–10}. Toll-like receptor 9 (TLR9) is known to recognize CpG DNA to trigger IFN production in plasmacytoid dendritic cells (pDCs), and Z-DNA binding protein 1 (ZBP1, also known as DAI) was recently shown to be able to stimulate IFN transcription, but was found to be largely redundant in studies using DAI-deficient cells and mice^{11–13}. Recently, a DNA receptor AIM2 was found to be important for ASC (also known as PYCARD)-dependent inflammasome mediated production of IL1 β , but was not required for type I IFN production^{14–18}. Thus, other innate signalling pathways that recognize intracellular non-CpG DNA species must exist to facilitate type I IFN production.

We previously demonstrated for the first time a role for STING (also referred to as TMEM173, MPYS and MITA), an endoplasmic reticulum (ER) resident transmembrane protein, in facilitating the production of type I IFN^{4,19,20}. To evaluate the importance of STING in mediating DNA-induced innate immune responses, we used wild type (+/+) or *Sting*^{-/-} low passage number mouse embryonic fibroblasts (MEFs) and compared the induction of type I IFN (IFN β) in response to a variety of DNA ligands. Our results indicated that

STING was essential for inducing IFN β in response to transfected viral DNA (adenovirus, Ad5; herpes simplex virus, HSV-1 and -2), purified *Escherichia coli* DNA, calf thymus (CT) DNA, and interferon stimulatory DNA (ISD; double-stranded 45-base-pair oligonucleotides lacking CpG sequences) (Fig. 1a). Complete abrogation of IFN β production was also observed after transfection of synthetic double-stranded DNA (poly(dG-dC)•poly(dC-dG), hereafter referred to as poly(dGC:dGC)) in *Sting*^{-/-} MEFs, and slight IFN β production was observed using poly(dAT:dAT), probably due to STING-independent, RIG-I (also known as DDX58)-dependent signalling^{21,22}. The loss of STING did not significantly affect poly(I:C)-mediated type I IFN production, which is largely governed by MDA5 (ref. 5). Concomitant analysis further indicated a marked reduction in IL6 production in *Sting*^{-/-} MEFs compared to controls after similar DNA transfections (Fig. 1a). ISD-mediated production of *Ifnb* and *Ifna* messenger RNA was not detectable in *Sting*^{-/-} MEFs compared to controls (Fig. 1b). Translocation of IRF3 or IRF7 was thus not observed in ISD-transfected *Sting*^{-/-} MEFs, indicating that STING probably functions in mediating intracellular-DNA-triggered IFN production upstream of TBK1 (Fig. 1c and Supplementary Fig. 1). NF- κ B signalling was also defective in *Sting*^{-/-} MEFs after exposure to transfected ISD (Supplementary Fig. 1). Given this, we next examined the importance of STING in facilitating intracellular-DNA-mediated production of type I IFN in antigen presenting cells. This analysis indicated that *Sting*^{-/-} macrophages transfected with ISD, or infected with the DNA pathogens HSV-1 or *Listeria monocytogenes*, were greatly defective in their ability to manufacture type I IFN (Fig. 1d). However, the cleavage of pro-caspase 1 and production of active IL1 β , which is AIM2-dependent, was unaffected by the loss of STING (Fig. 1e and Supplementary Fig. 1). Thus, STING functions independently of the AIM2 'inflammasome' pathway. Further analysis also indicated that STING was required for efficient DNA-mediated production of type I IFN in granulocyte-macrophage dendritic cells (GM-DCs), as well as pDCs (FLT3-ligand-induced dendritic cells, FLT3-DCs) (Fig. 1f, g). However, exogenous CpG DNA remained able to induce type I IFN in *Sting*^{-/-} FLT3-DCs compared to controls, indicating that TLR9 functions independently of the STING pathway (Fig. 1g). The induction of IL6 in response to intracellular DNA was also reduced in *Sting*^{-/-} macrophages (Supplementary Fig. 1). However, HSV-1 and CpG DNA remained able to induce IL6 in *Sting*^{-/-} macrophages, probably through TLR9-dependent signalling (Supplementary Fig. 1)¹¹. Furthermore, we noted that STING seemed to be essential for the production of type I IFN by cytomegalovirus (CMV), vaccinia virus (VV Δ E3L) and baculovirus (Supplementary Fig. 1). STING therefore seems critical for intracellular-DNA-mediated production of type I IFN in fibroblasts, macrophages, conventional dendritic cells as well as pDCs.

We next evaluated the *in vivo* importance of STING in facilitating effective host defence against select virus infection. Principally,

¹Department of Medicine and Sylvester Comprehensive Cancer Center, University of Miami Miller School of Medicine, Miami, Florida 33136, USA.

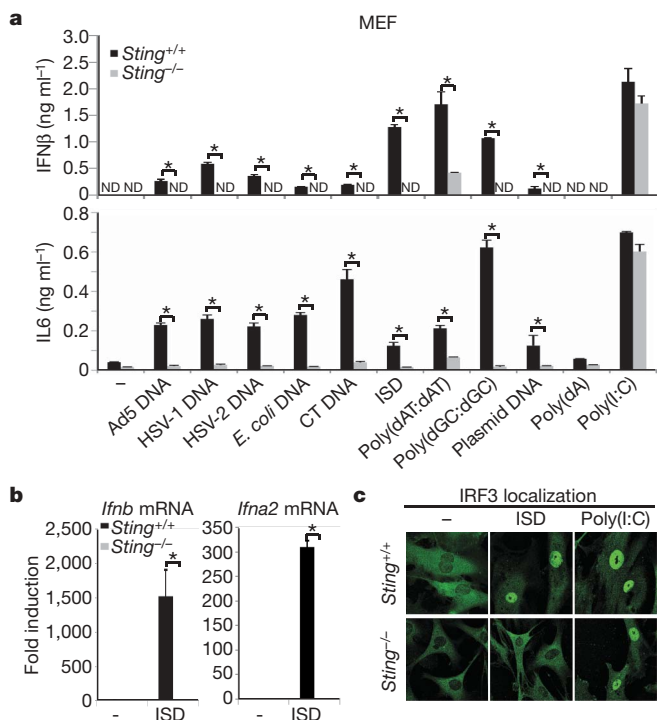
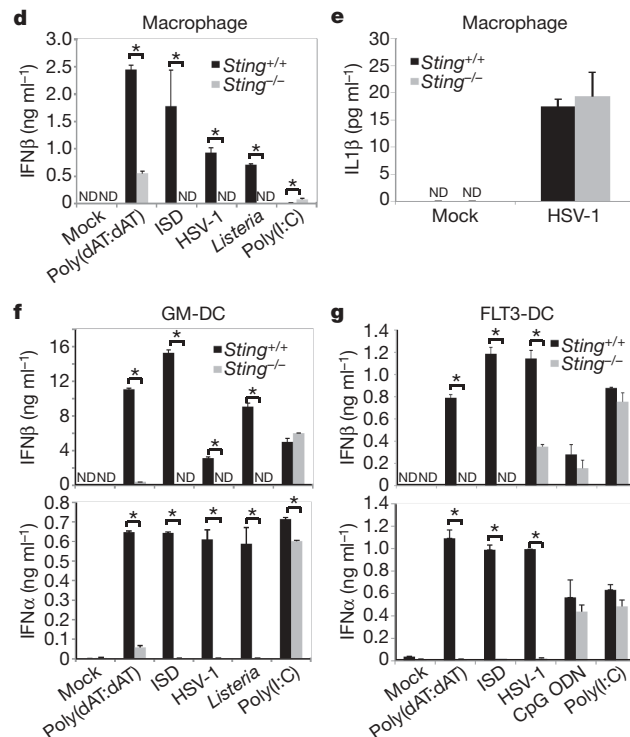


Figure 1 | STING is essential for intracellular DNA-mediated type I IFN production. **a**, MEFs were transfected with $1 \mu\text{g ml}^{-1}$ of DNA ligands (with Lipofectamine 2000) for 16 h, and IFN β or IL6 were measured. **b**, MEFs were transfected with ISD for 4 h and *Irfnb* or *Irfna2* mRNA levels were measured. **c**, MEFs treated as in **b** were stained with an antibody for IRF3 translocation. Original magnification, $\times 40$. **d**, Bone-marrow-derived macrophages were transfected with poly(dAT:dAT), poly(I:C) or ISD, or infected with HSV-1 (multiplicity of infection (m.o.i.) 10) or *Listeria* (m.o.i. 10) for 16 h, and

Sting^{-/-} or control mice were infected intravenously (i.v.) with HSV-1 and survival was monitored. The *Sting*-knockout mice died within 7 days of HSV-1 infection (Fig. 2a), whereas 80% of similarly infected wild-type mice survived. Significant amounts of HSV-1 were detected in the brain of infected *Sting*^{-/-} mice, but not in controls at 5 days after infection (Fig. 2b). Analysis of serum from the *Sting*^{-/-}-infected animals indicated a profound defect in the production of type I IFN at 6 h after infection, compared to infected control animals (Fig. 2c, d and Supplementary Fig. 2). RANTES and IL6 levels were similarly markedly reduced in *Sting*^{-/-} mice at the same time point (Fig. 2e, f). Moreover, *Sting*^{-/-} mice were found to be more sensitive to HSV-1 after intravaginal administration of HSV-1 (Supplementary Fig. 2). This data indicates that STING is necessary, *in vivo*, for the effective production of type I IFN and is essential for efficient protection against HSV-1 infection.

Because we had previously seen, *in vitro*, a defect in the ability of the negative-stranded virus vesicular stomatitis virus (VSV) to induce type I IFN in the absence of STING, we next examined the *in vivo* importance of STING in protecting against VSV-related disease⁴. We observed that *Sting*^{-/-} animals infected with VSV was also significantly sensitive to lethal infection compared to controls (Fig. 2g). Defects in type I IFN production were seen in *Sting*-knockout mice at early time points (6h), although less so at 24 h (Fig. 2h, i and Supplementary Fig. 2). Thus, STING is necessary for efficient, early induction of type I IFN production and is required for protection against infection with the negative-stranded virus VSV, possibly by regulating the RIG-I and IPS-1 (also known as MAVS, VISA and CARDIF) pathway^{4,6-10}.

We did not observe a significant requirement for STING in facilitating poly(I:C) or EMCV (encephalomyocarditis virus, a positive-stranded flavivirus)-mediated IFN transcription, indicating that



IFN β was measured. **e**, Macrophages were infected with HSV-1 for 16 h and IL1 β was measured. **f**, GM-colony stimulating factor (CSF)-induced dendritic cells (GM-DCs) were treated as in **d**, and IFN β or IFN α was measured after 16 h. **g**, FLT3-stimulated dendritic cells were treated as in **f** (exogenous CpG oligodeoxynucleotides (ODN) (1 μ g ml $^{-1}$) were also used). * P < 0.05, Student's t -test. Error bars indicate s.d. ND, not determined.

STING may not influence MDA5 function (Fig. 1a and Supplementary Fig. 2)⁴. However, it is known that some flaviviruses such as hepatitis C virus (HCV) can activate the RIG-I pathway, signalling which seems to be influenced by STING^{4,23}. In this regard, databank analysis indicated that the flaviviruses yellow fever virus (YFV) and Dengue virus encode a product NS4B that exhibits strong homology with the amino terminus of STING (amino acids 125–222). (Supplementary Fig. 3). This region was found to be critical for STING function (Supplementary Fig. 3). Various flaviviral NS4B products have been shown to localize to the ER of the cell and to suppress the induction of type I IFN, although the mechanisms remain unclear²⁴. Our analysis here indicates that that NS4B was able to inhibit STING activity, probably by direct association (Fig. 2j–l and Supplementary Fig. 3). Thus, STING may be targeted by certain viruses for suppression.

TBK1 has been shown to have an important role in mediating the adjuvant activity of DNA vaccines *in vivo*¹². TBK1 activation in response to plasmid DNA was found to occur in the absence of the DNA sensors TLR9 or DAI, indicating that other pathways exist to facilitate DNA-mediated immunization^{12,25}. To evaluate whether STING was involved in this signalling pathway, *Sting*^{-/-} or control mice were immunized with plasmid DNA encoding the ovalbumin gene. Although we noted normal B- and T-cell subsets in unstimulated *Sting*^{-/-} animals, after immunization *Sting*^{-/-} mice showed significantly less serum ovalbumin (OVA)-specific IgG compared to controls (Fig. 3a and Supplementary Fig. 4). Furthermore, spleen CD8⁺ T-cell frequency and IFN γ secretion was markedly reduced in *Sting*^{-/-} mice after immunization, compared to wild-type mice (Fig. 3b, c). Because immunoglobulin responses to OVA peptide were normal, these data emphasize that the STING-governed DNA sensor pathway is essential for efficient DNA-vaccine-induced T-cell

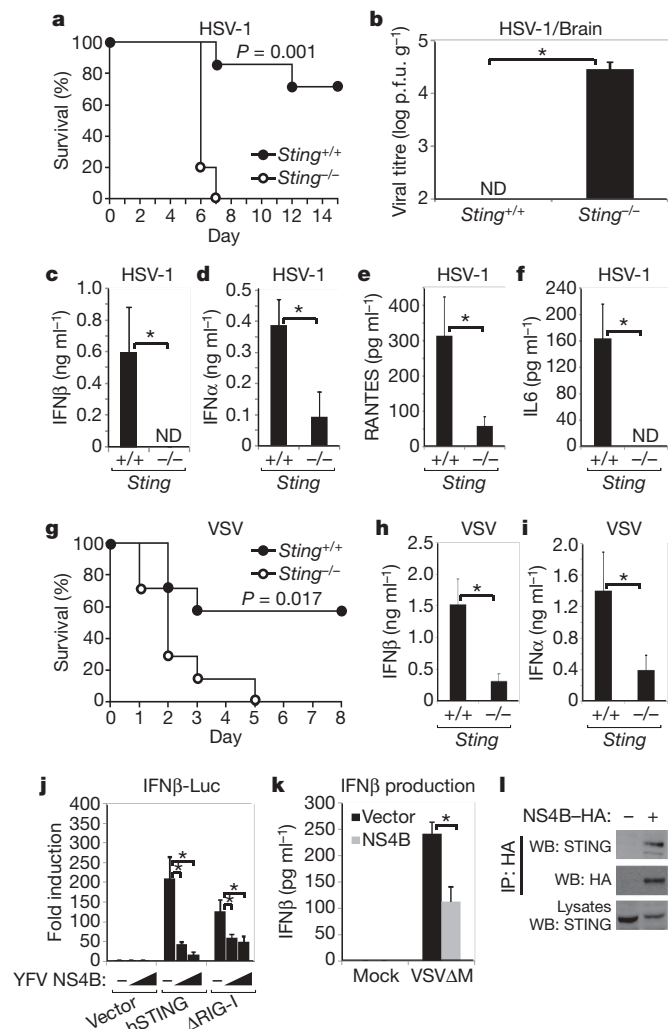


Figure 2 | STING is required for effective *in vivo* host defence. **a**, $Sting^{-/-}$ animals ($Sting^{-/-}$) or littermate controls ($Sting^{+/+}$) ($n = 7$; approximately 8-weeks-of-age) were infected with HSV-1 (1×10^7 i.v.) and survival was monitored. **b**, $Sting^{-/-}$ or control mice were infected with HSV-1 as in **a** and brains were retrieved after 5 days for HSV-1 plaque assays. p.f.u., plaque-forming units. **c**, **d**, Serum from animals ($n = 3$) infected with HSV-1 (1×10^7 i.v.) was analysed for IFN β (**c**) or IFN α (**d**) production after 6 h. **e**, **f**, Serum from animals infected as in **c** was analysed for RANTES (**e**) and IL6 (**f**) production. **g**, $Sting^{-/-}$ or control mice ($n = 6$) were infected with VSV (5×10^7 i.v.) and survival was monitored. **h**, **i**, Mice ($n = 3$) were treated as in **g** and IFN β (**h**) or IFN α (**i**) was measured after 6 h. **j**, Increasing amounts of YFV NS4B were co-transfected into 293T cells with human STING or the amino terminus of RIG-I (Δ RIG-I, residues 1–284) and transfected IFN β promoter-driven luciferase (IFN β -Luc) was measured after 36 h. **k**, Immortalized MEFs were transfected with YFV NS4B for 24 h, infected with VSV Δ M⁴ (m.o.i. 1) for 16 h, and IFN β was measured. **l**, 293 cells were transfected with NS4B-HA for 36 h and after immunoprecipitation (IP) with anti-haemagglutinin antibody, were analysed by western blot (WB) using anti-STING serum. $*P < 0.05$, Student's *t*-test. Error bars indicate s.d.

responses to antigen (Fig. 3 and Supplementary Fig. 4). Similar studies also indicated that STING had a key role in facilitating T-cell responses to the DNA virus vaccinia expressing ovalbumin (VV-OVA). Our data emphasizes the importance of STING in innate immune signalling processes required for DNA adjuvant activity (Fig. 3d).

We previously demonstrated that STING is an ER resident protein and member of the TRAP (translocon associated protein) complex that can associate with RIG-I and the mitochondrial innate immune signalling adaptor IPS-1 (refs 4, 26). Physical association of mitochondria

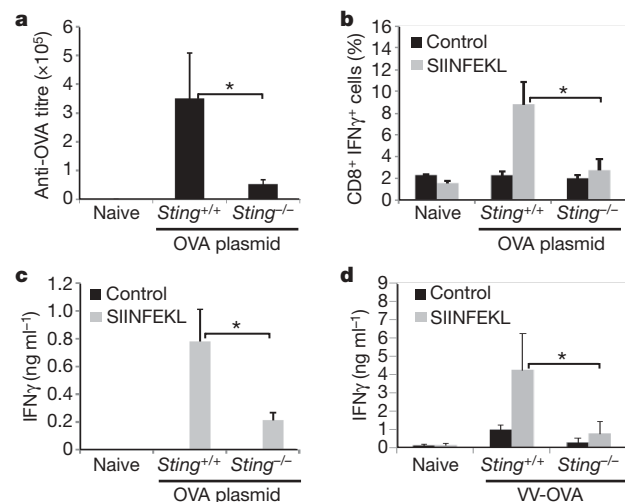


Figure 3 | STING is required for effective DNA-mediated adaptive immune responses. **a**, $Sting^{-/-}$ or control ($Sting^{+/+}$) mice ($n = 5$; approximately 8-weeks-of-age) were immunized twice (100 μ g i.m.) by electroporation with a DNA vaccine encoding ovalbumin. Serum was measured for anti-OVA IgG. **b**, **c**, Mice were treated as in **a** and spleen CD8 $^{+}$ IFN γ^{+} cells were measured by fluorescence-activated cell sorting (FACS; **b**), and anti-OVA-specific IFN γ production was measured by ELISA after stimulation of splenocytes using SIINFEKL peptide (**c**). **d**, $Sting^{-/-}$ mice or controls ($n = 4$; approximately 8-weeks-of-age) were infected with vaccinia expressing ovalbumin (VV-OVA; 5×10^6 i.v.) and spleen anti-OVA-specific IFN γ production was measured by ELISA. $*P < 0.05$, Student's *t*-test. Error bars indicate s.d. All experiments were repeated twice.

and the ER, referred to as mitochondria-associated ER membrane (MAM), is important for transmission of Ca^{2+} to the mitochondria and for oxidative metabolism²⁷. We thus examined whether STING could associate with MAMs. First, we reconstituted haemagglutinin (HA)-tagged STING into $Sting^{-/-}$ MEFs to follow endogenous STING localization using a haemagglutinin antibody. This analysis confirmed that STING is predominantly associated with the ER as determined by calreticulin marker co-staining (Fig. 4a). Mitotracker co-staining also indicated that STING may co-localize with mitochondria associated with the ER (Fig. 1b). The association of endogenous STING with the ER was also confirmed using anti-STING serum (Supplementary Fig. 5). Fractionation analysis subsequently demonstrated that STING is associated with microsomes, a complex of continuous membranes that comprise the ER, Golgi and transport vesicles (Fig. 4c). Endogenous STING was found to fractionate with MAMs and mitochondria fractions under non-stimulated conditions in MEFs (Fig. 4c). Calreticulin, known to be a chaperone involved in regulating the association of the ER and mitochondria, was observed to fractionate similarly²⁷. This data may indicate that STING could associate with IPS-1 by MAM interaction⁴. Interestingly, after HSV-1 infection, STING was shown to become predominantly associated only with microsome fractions (Fig. 4c). To clarify these observations, we infected STING-HA MEFs with HSV-1, or transfected these cells with stimulatory ISD or negative-control single-stranded DNA (ssDNA). These results indicated that in response to HSV-1 infection or ISD transfection, STING translocated from the ER and predominantly congregated to perinuclear, non-ER microsome compartments in the cell (Fig. 4d and Supplementary Figs 5 and 6). Brefeldin A, but not chloroquine, blocked STING trafficking, indicating that STING locates from the ER via the Golgi to vesicles in the perinuclear region (Supplementary Fig. 5). This trafficking, in response to intracellular DNA, was similarly observed for TBK1, which we have previously shown to associate with STING⁴ (Fig. 4e). Notably, in the absence of STING, TBK1 failed to relocate to perinuclear regions in response to ISD transfection (Supplementary Fig. 7).

We further observed that in the presence of DNA, STING mostly localized with the early endosome marker protein EEA1 and recycling

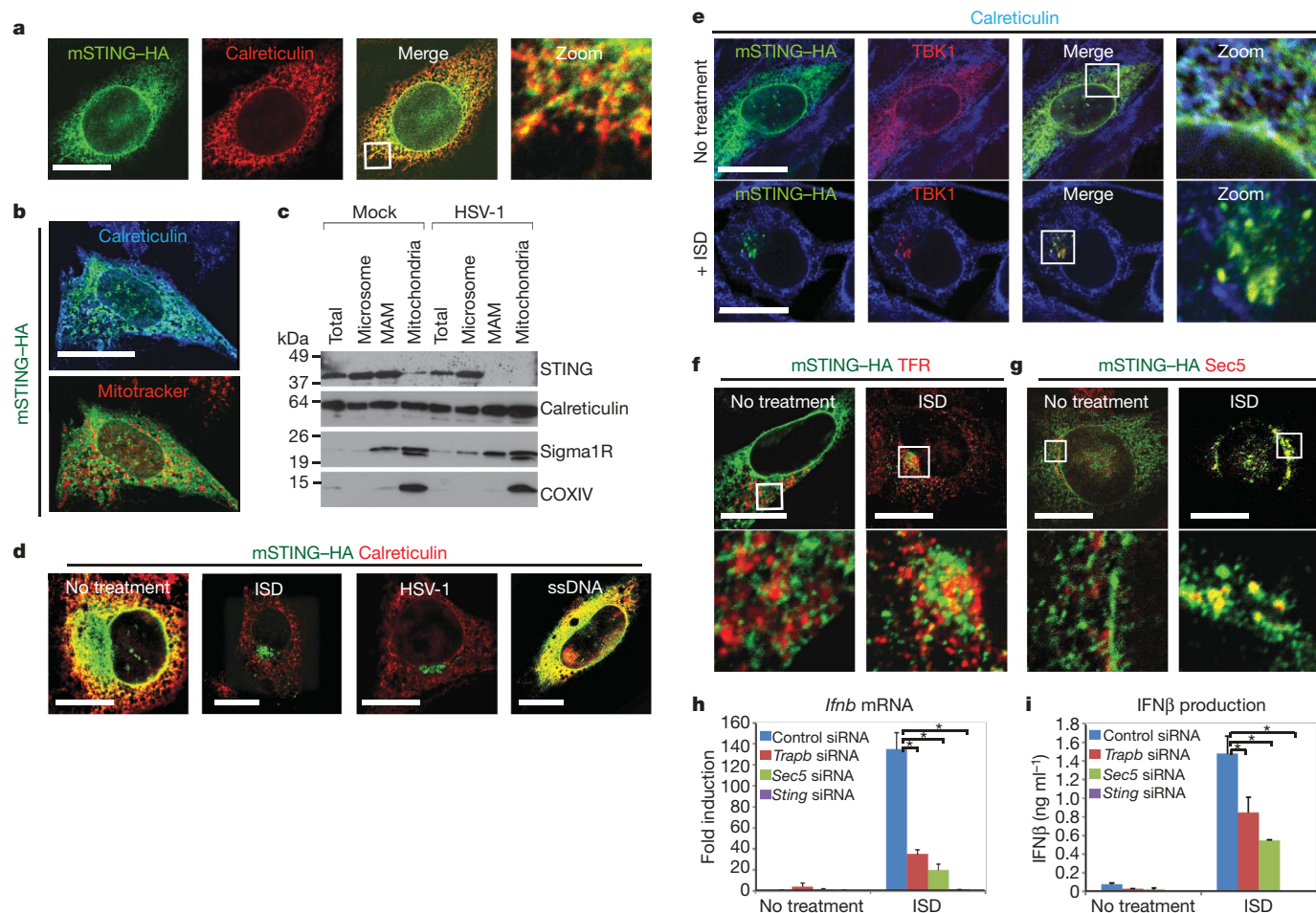


Figure 4 | **STING translocates from the ER to Sec5-containing vesicles.** **a**, *Sting*^{-/-} MEFs, stably reconstituted with haemagglutinin-tagged mouse STING (mSTING-HA) were stained using haemagglutinin (green) and a calreticulin (red) antibody. **b**, STING-HA MEFs were stained for STING-HA (green), calreticulin (blue) or Mitotracker (red) and three-dimensional reconstruction images were taken. **c**, Immunoblot analysis of fractionation experiments of uninfected or HSV-1-infected (m.o.i. 10; 4 h) MEFs. Endogenous STING was detected using an anti-STING antibody. Calreticulin detects ER, Sigma1R detects MAM, and COXIV detects mitochondria. **d**, Haemagglutinin (green) or calreticulin (red) staining of mSTING-HA MEFs after treatment with transfected ISD (1 $\mu\text{g ml}^{-1}$),

endosome marker transferrin receptor (TFR; Fig. 4f and Supplementary Fig. 6). TBK1 has also been demonstrated to associate with Sec5, a component of the exocyst 8 subunit complex that facilitates vesicular transport processes²⁸. After intracellular DNA stimulation, STING was found to strongly colocalize with Sec5, which has also been demonstrated to associate in perinuclear endosome compartments (Fig. 4g)²⁹. The RALB and Sec5 pathway has been previously shown to be required for efficient Sendai-virus-mediated type I IFN production²⁸. However, our data here indicates that STING and TBK1 complexes may traffic to endosome compartments to associate with Sec5/exocyst components and facilitate the production of type I IFN in response to intracellular DNA. To evaluate whether Sec5 also modulates the production of IFN β in response to ISD, we suppressed Sec5 production in normal MEFs using RNA interference (RNAi). This study indicated that in the absence of Sec5, ISD-mediated IFN production was significantly impaired (Fig. 4h, i). A similar effect was observed after knockdown of *Trapb* (also known as *Ssr2*) and *Sec61b*, components of the TRAP complex (Fig. 4h, i and Supplementary Fig. 8). Our data thus indicates that intracellular DNA may induce STING to complex with TBK1 and traffic to Sec5-containing endosome compartments—events that facilitate the production of type I IFN.

transfected ssDNA (1 $\mu\text{g ml}^{-1}$) or HSV-1 infection as in **c**. **e**, mSTING-HA MEFs were transfected with or without ISD and cells were stained with haemagglutinin (green), calreticulin (blue) and a TBK1 (red) antibody. **f**, mSTING-HA MEFs were transfected as in **e** and stained with haemagglutinin (green) and a TFR (red) antibody. **g**, mSTING-HA MEFs were transfected as in **e** and stained with haemagglutinin (green) and a Sec5 antibody (red). **h**, **i**, MEFs were treated with RNAi to *Trapb*, *Sting* or *Sec5* for 72 h and transfected with ISD. IFN β mRNA and protein were measured at 4 and 16 h, respectively. **P* < 0.05, Student's *t*-test. Error bars indicate s.d. Scale bars, 10 μm .

In conclusion, we demonstrate that STING is essential for the recognition of intracellular DNA and efficient production of type I IFN in all cell types examined. Loss of STING renders mice susceptible to lethal DNA virus infection (HSV-1). However STING also facilitates host defence responses to negative-stranded viruses such as VSV, plausibly through RIG-I and IPS-1–MAM translocon interactions. Although STING-independent, VSV-mediated type I IFN-induction pathways clearly exist, they do not seem to be sufficient on their own to protect mice against lethal VSV infection. We conclude that in response to intracellular DNA, STING and TBK1 complexes traffic to endosomal compartments to associate with exocyst components including Sec5, resulting in the induction of type I IFN.

METHODS SUMMARY

Details of mice, cells, viruses, plasmids, antibodies and reagents are given in the Methods. ELISA kits were obtained from following sources: murine IFN β and IFN α (PBL), murine IL6 (R&D systems or Quansys Biosciences), murine IL1 β and IFN γ (R&D systems), active NF- κ B p65 (Active Motif) murine RANTES (Quansys Biosciences).

DNA vaccine. Mice were immunized with a plasmid encoding OVA by intramuscular (i.m.) electroporation (100 μg per mouse). The booster immunization was given within 4 weeks of the primary immunization.

Measurement of OVA-specific immune response. Spleen cells were extracted 2 weeks after the second immunization and stimulated with synthetic peptide for OVA (H-2Kb SIINFEKL, Proimmune) at $10 \mu\text{g ml}^{-1}$. After 3 days, the cell culture supernatants were collected and analysed for the IFN γ titre by ELISA (R&D systems). For intracellular IFN γ staining, stimulated splenocytes were stained using FITC-labelled anti-CD8 antibody (BD). The serum anti-OVA antibody titre was measured by ELISA. Further details are given in the Methods.

Confocal microscopy. For localization of Sec5 and LAMP1, cells grown on coverslips were fixed in 80%/20% methanol/acetone at -20°C for 5 min. For EEA1 staining, cells were fixed with 4% paraformaldehyde in PBS for 15 min at 37°C , and were permeabilized in 0.2% Triton X-100. For staining of other proteins, cells were fixed with 4% formaldehyde in DMEM for 15 min at 37°C , and were permeabilized in 0.2% Triton X-100. For mitochondria staining, living cells were incubated with 300 nM of Mito Tracker Red (Invitrogen) for 45 min at 37°C .

RNA interference. Chemically synthesized 21-nucleotide short interfering RNA (siRNA) duplexes were obtained from Dharmacon, Inc. The sequences of each siRNA oligonucleotide used in this study are given in the Methods. MEFs were transfected using an Amaxa nucleofactor apparatus (program A-023) and Amaxa MEF nucleofactor kit 1 according to the manufacturer's instructions.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 2 August; accepted 3 September 2009.

Published online 23 September 2009.

- Palm, N. W. & Medzhitov, R. Pattern recognition receptors and control of adaptive immunity. *Immunol. Rev.* **227**, 221–233 (2009).
- Takeuchi, O. & Akira, S. Innate immunity to virus infection. *Immunol. Rev.* **227**, 75–86 (2009).
- Beutler, B. A. TLRs and innate immunity. *Blood* **113**, 1399–1407 (2009).
- Ishikawa, H. & Barber, G. N. STING is an endoplasmic reticulum adaptor that facilitates innate immune signalling. *Nature* **455**, 674–678 (2008).
- Kato, H. *et al.* Differential roles of MDA5 and RIG-I helicases in the recognition of RNA viruses. *Nature* **441**, 101–105 (2006).
- Yoneyama, M. *et al.* The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses. *Nature Immunol.* **5**, 730–737 (2004).
- Kawai, T. *et al.* IPS-1, an adaptor triggering RIG-I- and Mda5-mediated type I interferon induction. *Nature Immunol.* **6**, 981–988 (2005).
- Seth, R. B., Sun, L., Ea, C. K. & Chen, Z. J. Identification and characterization of MAVS, a mitochondrial antiviral signaling protein that activates NF- κB and IRF 3. *Cell* **122**, 669–682 (2005).
- Meylan, E. *et al.* Cardif is an adaptor protein in the RIG-I antiviral pathway and is targeted by hepatitis C virus. *Nature* **437**, 1167–1172 (2005).
- Xu, L. G. *et al.* VISA is an adapter protein required for virus-triggered IFN- β signaling. *Mol. Cell* **19**, 727–740 (2005).
- Bauer, S., Pigisch, S., Hangel, D., Kaufmann, A. & Hamm, S. Recognition of nucleic acid and nucleic acid analogs by Toll-like receptors 7, 8 and 9. *Immunobiology* **213**, 315–328 (2008).
- Ishii, K. J. *et al.* TANK-binding kinase-1 delineates innate and adaptive immune responses to DNA vaccines. *Nature* **451**, 725–729 (2008).
- Takaoka, A. *et al.* DAI (DLM-1/ZBP1) is a cytosolic DNA sensor and an activator of innate immune response. *Nature* **448**, 501–505 (2007).
- Muruve, D. A. *et al.* The inflammasome recognizes cytosolic microbial and host DNA and triggers an innate immune response. *Nature* **452**, 103–107 (2008).
- Roberts, T. L. *et al.* HIN-200 proteins regulate caspase activation in response to foreign cytoplasmic DNA. *Science* **323**, 1057–1060 (2009).
- Hornung, V. *et al.* AIM2 recognizes cytosolic dsDNA and forms a caspase-1-activating inflammasome with ASC. *Nature* **458**, 514–518 (2009).
- Fernandes-Alnemri, T., Yu, J. W., Datta, P., Wu, J. & Alnemri, E. S. AIM2 activates the inflammasome and cell death in response to cytoplasmic DNA. *Nature* **458**, 509–513 (2009).
- Bürkstümm, T. *et al.* An orthogonal proteomic-genomic screen identifies AIM2 as a cytoplasmic DNA sensor for the inflammasome. *Nature Immunol.* **10**, 266–272 (2009).
- Jin, L. *et al.* MPYS, a novel membrane tetraspanner, is associated with major histocompatibility complex class II and mediates transduction of apoptotic signals. *Mol. Cell. Biol.* **28**, 5014–5026 (2008).
- Zhong, B. *et al.* The adaptor protein MITA links virus-sensing receptors to IRF3 transcription factor activation. *Immunity* **29**, 538–550 (2008).
- Ablasser, A. *et al.* RIG-I-dependent sensing of poly(dA:dT) through the induction of an RNA polymerase III-transcribed RNA intermediate. *Nature Immunol.* doi:10.1038/ni.1779 (16 July 2009).
- Chiu, Y. H., Macmillan, J. B. & Chen, Z. J. RNA polymerase III detects cytosolic dna and induces type I interferons through the RIG-I pathway. *Cell* **138**, 576–591 (2009).
- Saito, T., Owen, D. M., Jiang, F., Marcotrigiano, J. & Gale, M. Jr. Innate immunity induced by composition-dependent RIG-I recognition of hepatitis C virus RNA. *Nature* **454**, 523–527 (2008).
- Muñoz-Jordan, J. L. *et al.* Inhibition of α/β interferon signaling by the NS4B protein of flaviviruses. *J. Virol.* **79**, 8004–8013 (2005).
- Spies, B. *et al.* Vaccination with plasmid DNA activates dendritic cells via Toll-like receptor 9 (TLR9) but functions in TLR9-deficient mice. *J. Immunol.* **171**, 5908–5912 (2003).
- Ménétret, J. F. *et al.* Single copies of Sec61 and TRAP associate with a nontranslating mammalian ribosome. *Structure* **16**, 1126–1137 (2008).
- Hayashi, T., Rizzuto, R., Hajnoczky, G. & Su, T. P. MAM: more than just a housekeeper. *Trends Cell Biol.* **19**, 81–88 (2009).
- Chien, Y. *et al.* RalB GTPase-mediated activation of the I κB family kinase TBK1 couples innate immune signaling to tumor cell survival. *Cell* **127**, 157–170 (2006).
- Spiczka, K. S. & Yeaman, C. Ral-regulated interaction between Sec5 and paxillin targets Exocyst to focal complexes during cell migration. *J. Cell Sci.* **121**, 2880–2891 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Yewdell for VV-OVA, B. Jacobs for VV ΔE3L , K. Frueh for HCMV, M. Kobayashi for baculovirus, H. Horiuchi for the Sec5 antibody, Y. C. Weh for *Tbk1*-knockout MEFs, and S. Nagata, T. Maniatis, J. Hiscott and N. Reich for plasmid constructs. This work was supported by NIH grant AI079336.

Author Contributions H.I. and G.N.B. designed the research and analysed the data. H.I. performed most experiments. Z.M. performed experiments related to YFV NS4B, carried out exocyst RNAi studies and helped with experiments. G.N.B. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to G.N.B. (gbarber@med.miami.edu).

METHODS

Mice, cells, viruses and reagents. *Sting*-knockout mice on a 129SvEv × C57BL/6J background have been described previously⁴. MEFs, bone-marrow-derived macrophages and GM-DCs were prepared as described previously⁴. To prepare FLT3-DCs, bone marrow cells were cultured in RPMI 1640 medium supplemented with 10% FBS, 50 μM 2-mercaptoethanol, 10 mM HEPES, pH 7.4, and 100 ng ml⁻¹ human FLT3 ligand (Peprotech) for 8 days. 293T cells were obtained from the American Type Culture Collection (ATCC) and were maintained in DMEM medium supplemented with 10% FBS. VSV (Indiana strain), VSVΔM and EMCV were described previously⁴. HSV-1 (KOS strain) and *Listeria monocytogenes* (10403 serotype) were obtained from ATCC. Vaccinia virus encoding chicken ovalbumin (VV-OVA), vaccinia virus E3L deletion mutant (VVΔE3L), human cytomegalovirus (AD169 strain), and baculovirus (*Autographa californica* M nucleopolyhedrovirus) were gifts from J. Yewdell, B. Jacobs, K. Frueh and M. Kobayashi, respectively. Genomic DNA was obtained from following sources: HSV-1, HSV-2, adenovirus type 5 (ATCC); *E. coli*, and calf thymus (Sigma). Poly(dAT:dAT) and poly(I:C) were obtained from Amersham Biosciences. Poly(dGC:dGC) and poly(dA) were obtained from Sigma. CpG ODN (ODN 1585) was obtained from Invivogen. For stimulation of cells, genomic DNA, polydeoxynucleotides or poly(I:C) were mixed with Lipofectamine 2000 (Invitrogen) at a ratio of 1:1 (v/w), and then added to cells at a final concentration of 1 μg ml⁻¹. LPS was obtained from Invivogen. Brefeldin A and chloroquine were obtained from Sigma.

Plasmids. YFV NS4B sequence was amplified by PCR using pYFM5.2 encoding the complete YFV-17D sequence as a template, and was cloned into a pcDNA3 (Invitrogen) plasmid to generate carboxy-terminally haemagglutinin-tagged expression construct. C-terminally haemagglutinin-tagged STINGΔSP (Δ1–36 amino acids) and STINGΔTM5 (Δ153–173 amino acids) were amplified by PCR and cloned into a pcDNA3 plasmid. The expression plasmid containing chicken ovalbumin (OVA) complementary DNA was constructed by cloning of PCR-amplified OVA cDNA into pcDNA3. Expression plasmids encoding haemagglutinin-tagged murine STING (mSTING-HA), Flag-tagged ΔRIG-I (amino acids 1–284), ΔMDA5 (amino acids 1–349) and IRF-7 were described previously⁴. p110-Luc (IFNβ-Luc) was obtained from T. Maniatis. pUNO-hsaIRF3 (IRF3SA) and pUNO-hsaIRF7Δ (IRF7SA) were obtained from Invivogen. pCMV-SPORT6 containing murine DAI was obtained from Open Biosystems.

Primers. The following primers were used for cloning: YFV NS4B forward, 5'-GGGGTACCATGAACGAGCTAGGCATGCTGGAG-3'; YFV NS4B reverse, 5'-CCGCTCGAGCCGCGCTCCAGTTTTCATCTTC-3'; STINGΔSP forward, 5'-CCCAAGCTTGCGCCACCATGCTAGGAGAGCCACAGAGCAC-3'; STINGΔSP reverse, 5'-CCGCTCGAGAGAGAAATCCGTGCGGAGAG-3'; OVA forward, 5'-ATGGCTCCATCGGCGCAGCAA-3'; OVA reverse, 5'-TTAAGGGGAAACACATCTGCC-3'.

Antibodies and ELISA. Rabbit polyclonal antibody against STING was described previously⁴. The antibody against STING-C was generated by immunizing rabbit with recombinant glutathione S-transferase (GST)-hSTING-C (amino acids 173–379) produced in *E. coli*. Rabbit polyclonal antibody against Sec5 was a gift from H. Horiuchi. Other antibodies were obtained from following sources: caspase-1 p10 (Santa Cruz Biotechnology), calreticulin (ab14234; Abcam), Signal receptor (ab53852; Abcam), TBK1 (EP611Y, Abcam), COXIV (ab16056, Abcam), rabbit polyclonal HA (ab9110; Abcam), transferrin receptor (H68.4; Invitrogen), mouse monoclonal haemagglutinin (Sigma), Flag (M2; Sigma), IRF3 (ZM3; Zymed), TGN46 (ab16059; Abcam), giantin (ab24586; Abcam), EEA1 (no.2441; Cell Signaling), LAMP1 (NB120; Novus Biologicals) and Sec61β (Upstate). ELISA kits were obtained from following sources: murine IFNβ and IFNα (PBL), murine IL6 (R&D systems or Quansys Biosciences), murine IL1β and IFNγ (R&D systems), active NF-κB p65 (Active Motif), and murine RANTES (Quansys Biosciences).

Real-time PCR. Fluorescence real-time PCR analysis was performed using a LightCycler 2.0 instrument (Roche Molecular Biochemicals) and the following TaqMan Gene Expression Assays (Applied Biosystems): IFNβ (Mm00439546_s1), IFNα2 (Mm00833961_s1) and TRAPβ (Mm00481383_m1). Relative amounts of mRNA were normalized to the 18S ribosomal RNA levels in each sample.

Reporter analysis. 293T cells seeded on 24-well plates were transiently transfected with 50 ng of the luciferase reporter plasmid together with a total of 600 ng of various expression plasmids or empty control plasmids. As an internal control, 10 ng pRL-TK was transfected simultaneously. Then, 24 or 36 h later, the luciferase activity in the total cell lysate was measured.

DNA vaccine. Mice were immunized with a plasmid encoding OVA by i.m. electroporation (100 μg per mouse). The booster immunization was given within 4 weeks of the primary immunization.

Measurement of OVA-specific immune response. Spleens were extracted 2 weeks after the second immunization and 5 × 10⁵ spleen cells were seeded on 96-well plates and then stimulated with synthetic peptide for OVA (H-2Kb SIINFEKL, Proimmune) at 10 μg ml⁻¹. After 3 days, the cell culture supernatants were collected and analysed for the IFNγ titre by ELISA (R&D systems). For intracellular IFNγ staining, stimulated splenocytes were stained using FITC-labelled anti-CD8 antibody (BD). After washing, cells were fixed and permeabilized. Then cells were stained using phycoerythrin (PE)-labelled anti-IFNγ antibody (BD). Flow cytometric analysis was performed on a FACScaliber instrument (BD). The serum anti-OVA antibody titre was measured by ELISA. In brief, 96-well plates were coated with an OVA protein at 1 μg ml⁻¹ and then blocked with PBS containing 5% skimmed milk. Plates were washed and overlaid with serially diluted serum for 1 h at room temperature. After washing, antibodies were detected using goat anti-mouse IgG conjugated to horseradish peroxidase (Jackson Immuno Research). After further washing, the plates were stained using 3,3',5,5'-tetramethylbenzidine (TMB, Sigma) as a substrate. The reaction was stopped with 1 M H₂SO₄ and the absorbance was measured. Antibody titres were expressed as the reciprocal of the endpoint dilution after background subtraction.

Fractionation. MAM, mitochondria and microsomes were isolated from *Sting*^{-/-} MEFs stably transfected with mSTING-HA plasmid as previously described³⁰. In brief, cells were washed in PBS and pelleted by centrifugation at 1,000g for 10 min. The pellet was resuspended in sucrose homogenization buffer (0.25 M sucrose, 10 mM HEPES, pH 7.4), and cells were lysed by using a dounce homogenizer. Lysed cells were centrifuged at 500g for 10 min, and the supernatant was collected. The supernatant was then centrifuged at 10,300g for 10 min to separate the crude microsomal (microsome and cytosol) from the crude mitochondrial (MAM and mitochondria) fraction, and the crude microsomal fraction (supernatant) was subjected to ultracentrifugation at 100,000g for 60 min. The crude mitochondrial fraction (pellet) was resuspended in ice-cold mannitol buffer A (0.25 M mannitol, 5 mM HEPES, 0.5 mM EDTA) and layered on top of a 30% Percoll in mannitol buffer B (0.225 M mannitol, 25 mM HEPES, 1 mM EDTA). Mitochondria and MAM fractions were separated by ultracentrifugation at 95,000g for 65 min. Both isolated fractions were diluted with mannitol buffer B and centrifuged at 6,300g for 10 min. The supernatant of MAM centrifugation was further separated by centrifugation at 100,000g for 60 min and the pellet was used for the MAM fraction, whereas the pellet of the mitochondria centrifugation was used as the mitochondria fraction. All of the fractions were resuspended in mannitol buffer B.

Confocal microscopy. For localization of Sec5 and LAMP1, cells grown on coverslips were fixed in 80%/20% methanol/acetone at -20 °C for 5 min. For EEA1 staining, cells were fixed with 4% paraformaldehyde in PBS for 15 min at 37 °C, and were permeabilized in 0.2% Triton X-100. For staining of other proteins, cells were fixed by 4% formaldehyde in DMEM for 15 min at 37 °C, and were permeabilized by 0.2% Triton X-100. For mitochondria staining, living cells were incubated with 300 nM of Mito Tracker Red (Invitrogen) for 45 min at 37 °C. Fixed and permeabilized cells were pre-incubated with 0.1% BSA in PBS and were incubated with primary antibodies. Cells were then incubated with secondary antibodies conjugated with FITC, Cy3 or Cy5 (Sigma).

RNA interference. Chemically synthesized 21-nucleotide siRNA duplexes were obtained from Dharmacon, Inc. The sequences of each siRNA oligonucleotide used in this study are as follows: murine *Trapb* siRNA, 5'-UGAAAGAGAGGACGGUUAUU-3'; murine *Sec5* siRNA, 5'-AGAAGUAUUAGGUCGGA-3', 5'-UCAACGUACUUCACGGAUU-3', 5'-CAGCAGAGAUUACACGUA-3', 5'-GUGAGUGGCUUGCGCAGUA-3'; murine *Sting* siRNA, 5'-CCAACAGCGUCUACGAGA-3'; murine *Sec61b* siRNA, 5'-GCAAGUACACGCGAUCAUA-3', 5'-CAUCGCGUGUUAUUUAUG-3', 5'-CCACUGUUCGGCAGAGAA-3', 5'-GGCGAUUCUACACGGAAGA-3'. Control siRNA was obtained from Dharmacon (D-001206-01-80). MEFs were transfected by using an Amaxa nucleofactor apparatus (program A-023) and Amaxa MEF nucleofactor kit 1 according to the manufacturer's instructions. L929 cells were transfected using Lipofectamine RNAiMAX (Invitrogen). At 72 h after transfection, cells were used for further experiments.

Statistics. Student's *t*-test was used to analyse data.

30. Mavinakere, M. S., Williamson, C. D., Goldmacher, V. S. & Colberg-Poley, A. M. Processing of human cytomegalovirus UL37 mutant glycoproteins in the endoplasmic reticulum lumen prior to mitochondrial importation. *J. Virol.* **80**, 6771–6783 (2006).

Prohibitin couples diapause signalling to mitochondrial metabolism during ageing in *C. elegans*

Marta Artal-Sanz¹ & Nektarios Tavernarakis¹

Marked alterations in cellular energy metabolism are a universal hallmark of the ageing process¹. The biogenesis and function of mitochondria, the energy-generating organelles in eukaryotic cells, are primary longevity determinants. Genetic or pharmacological manipulations of mitochondrial activity profoundly affect the lifespan of diverse organisms². However, the molecular mechanisms regulating mitochondrial biogenesis and energy metabolism during ageing are poorly understood. Prohibitins are ubiquitous, evolutionarily conserved proteins, which form a ring-like, high-molecular-mass complex at the inner membrane of mitochondria³. Here, we show that the mitochondrial prohibitin complex promotes longevity by modulating mitochondrial function and fat metabolism in the nematode *Caenorhabditis elegans*. We found that prohibitin deficiency shortens the lifespan of otherwise wild-type animals. Notably, knockdown of prohibitin promotes longevity in diapause mutants or under conditions of dietary restriction. In addition, prohibitin deficiency extends the lifespan of animals with compromised mitochondrial function or fat metabolism. Depletion of prohibitin influences ATP levels, animal fat content and mitochondrial proliferation in a genetic-background- and age-specific manner. Together, these findings reveal a novel mechanism regulating mitochondrial biogenesis and function, with opposing effects on energy metabolism, fat utilization and ageing in *C. elegans*. Prohibitin may have a similar key role in modulating energy metabolism during ageing in mammals.

The mitochondrial prohibitin complex comprises two subunits (PHB-1 and PHB-2) that assemble at the inner mitochondrial membrane⁴. Prohibitins have been implicated in several important cellular processes such as mitochondrial biogenesis and function, signalling, transcriptional control, cell death and replicative senescence. In addition, prohibitins have been associated with various types of cancer (reviewed in ref. 5). Little is known about the role of prohibitin in chronological ageing. We examined the requirement for prohibitin during ageing in *C. elegans*. Prohibitin genes are widely expressed in animal tissues throughout development and during adulthood. Green fluorescent protein (GFP)-tagged PHB-1 and -2 co-localize in mitochondria (Supplementary Fig. 1; ref. 6). Elimination of either PHB-1 or PHB-2 by RNA interference (RNAi; Supplementary Fig. 2) disrupts the mitochondrial prohibitin complex and causes early embryonic lethality⁶. Homozygous mutants harbouring a null *phb-1* allele become gametogenesis-defective sterile adults due to maternal effect (see Methods).

Post-embryonic RNAi knockdown of either *phb-1* or *phb-2* shortens the lifespan of otherwise wild-type worms (Fig. 1a and Supplementary Table 1). In sharp contrast, prohibitin deficiency markedly extends the lifespan of long-lived *daf-2* mutants (Fig. 1b and Supplementary Table 1). The insulin/insulin-like growth factor (IGF) receptor DAF-2 is a component of a signalling pathway regulating diapause entry (dauer larva formation)⁷. Longevity conferred by *daf-2* mutations requires

the DAF-16/FOXO transcription factor (reviewed in ref. 8). Loss of DAF-16 fully suppresses the exceptional longevity of prohibitin-depleted, *daf-2* mutants (Fig. 1c and Supplementary Table 1).

The transforming growth factor- β (TGF- β) signal transduction pathway also controls diapause and ageing⁹. The *daf-7* and *daf-4* genes encode a TGF- β homologue and the type II, transmembrane TGF- β receptor serine/threonine kinase, respectively^{10,11}. Prohibitin depletion

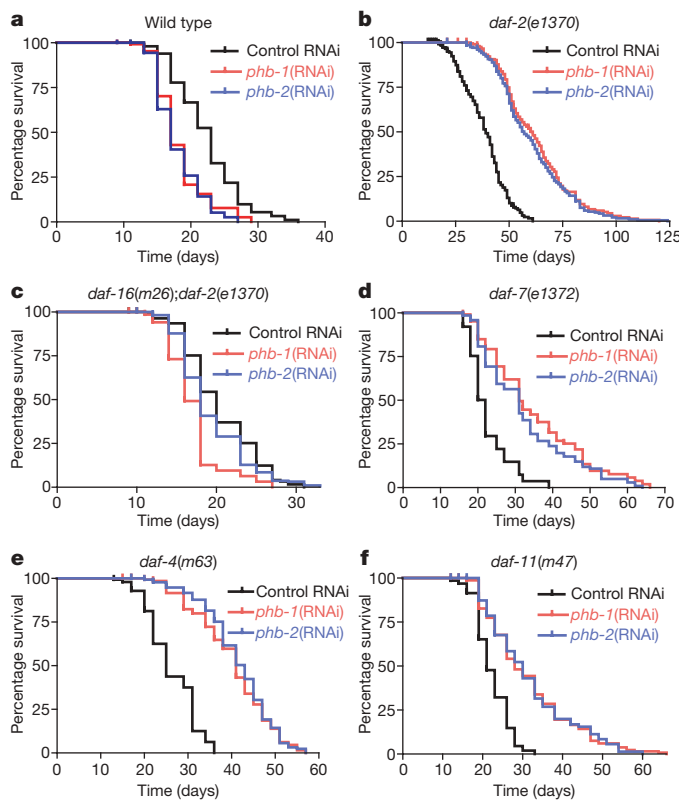


Figure 1 | Prohibitin deficiency markedly extends the lifespan of dauer-defective *C. elegans* mutants while shortening the lifespan of otherwise wild-type animals. The percentage of animals remaining alive is plotted against animal age. Assays were carried out at 20 °C. Combined lifespan data from independent experiments are given in Supplementary Table 1. **a**, Depletion of either PHB-1 or PHB-2 by RNAi in wild-type (N2) animals shortens lifespan. **b**, Prohibitin knockdown further extends the lifespan of long-lived, insulin signalling-defective *daf-2*(e1370) mutant animals. **c**, The longevity of prohibitin-depleted, *daf-2*(e1370) mutants is dependent on the transcription factor DAF-16/FOXO. **d**, Knockdown of either the *phb-1* or the *phb-2* gene extends the lifespan of *daf-7*(e1372) mutant animals, defective in TGF- β signalling. **e**, Survival curves of dauer-defective *daf-4* mutants subjected to RNAi with either *phb-1* or *phb-2*. **f**, Survival curves of dauer-defective *daf-11* mutants subjected to RNAi with either *phb-1* or *phb-2*.

¹Institute of Molecular Biology and Biotechnology, Foundation for Research and Technology, Heraklion 71110, Crete, Greece.

extends the lifespan of both *daf-7* and *daf-4* mutant animals (Fig. 1d, e and Supplementary Table 1). Furthermore, knockdown of *phb-1* or *phb-2* extends the lifespan of animals carrying a lesion in the *daf-11* gene, which encodes a transmembrane guanylate cyclase that functions via both the insulin/IGF and the TGF- β pathways to modulate dauer formation¹² (Fig. 1f and Supplementary Table 1). Thus, depending on the genetic background, prohibitin function has opposing effects on *C. elegans* ageing. Although depletion of prohibitin compromises survival in wild-type animals, it substantially extends the lifespan of mutants defective in either of the two diapause signalling pathways.

Both PHB-1 and PHB-2 proteins localize in mitochondria, where they form a high-molecular-mass complex (Supplementary Fig. 1; ref. 6). We investigated the role of prohibitin during ageing in animals carrying mutations that affect the mitochondrial electron transport chain. Knockdown of *phb-1* or *phb-2* extends the lifespan of *gas-1* mutants (Fig. 2a and Supplementary Table 1). The *gas-1* gene encodes a homologue of the 49-kDa iron-sulphur subunit of the mitochondrial electron transport chain complex I. Similarly, prohibitin depletion extends the lifespan of nematodes with lesions in the *mev-1* and *isp-1* genes (Fig. 2b, c and Supplementary Table 1). *mev-1* and *isp-1* encode the succinate dehydrogenase cytochrome *b*, a component of

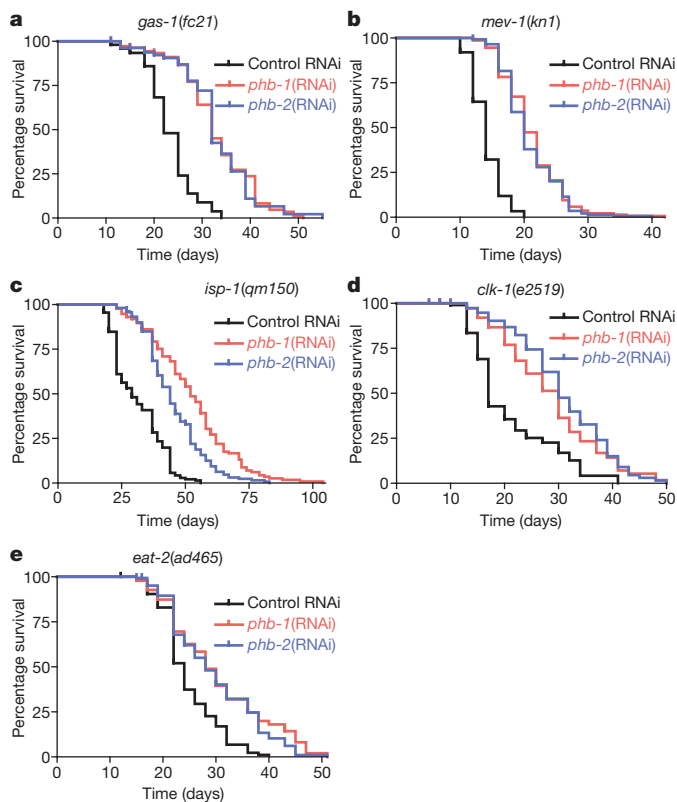


Figure 2 | Prohibitin deficiency further extends the lifespan of mitochondrial and dietary-restricted *C. elegans* mutants. Survival curves of mutant animal populations subjected to *phb-1* or *phb-2* RNAi are shown. **a**, Knockdown of prohibitin in mutants carrying a lesion in the *gas-1* gene, which encodes a homologue of the 49-kDa iron-sulphur protein fraction subunit of the mitochondrial NADH:ubiquinone-oxidoreductase, a component of the mitochondrial electron transport chain complex I. **b**, Knock-down of prohibitin in *mev-1* mutants, which lack succinate dehydrogenase cytochrome *b*, a component of the mitochondrial electron transport chain complex II. **c**, Knock-down of prohibitin in *isp-1* mutants, deficient for the Rieske iron-sulphur protein (ISP), a subunit of the mitochondrial electron transport chain complex III. **d**, Knock-down of prohibitin in long-lived animals carrying a mutation in the *clk-1* gene encoding a mitochondrial ubiquinone biosynthesis enzyme. **e**, Knock-down of prohibitin in long-lived, dietary-restricted *eat-2* mutants. The percentage of animals remaining alive is plotted against animal age. Lifespan values are given in Supplementary Table 1; assays were carried out at 20 °C.

complex II, and the Rieske iron-sulphur protein, a subunit of complex III, respectively. In addition, prohibitin knockdown extends the lifespan of *clk-1* mutant animals (Fig. 2d and Supplementary Table 1), which are defective in the biosynthesis of ubiquinone, an essential component of the electron transport chain. Hence, reduced prohibitin activity promotes survival of animals with compromised mitochondrial function. Energy metabolism in mitochondria is also affected by dietary restriction. We find that prohibitin deficiency improves the survival of dietary-restricted *eat-2* mutants (Fig. 2e and Supplementary Table 1).

To gain insight into the mechanism underlying the distinctive effects of prohibitin on ageing, we performed a temporal analysis of ATP levels during ageing (at day 3, 10 and 15 of adulthood) in wild-type animals and in *daf-2* and *daf-7* diapause mutants. We found that ATP levels are higher in diapause mutants compared to age-matched wild-type control animals. Knockdown of prohibitin specifically increased the levels of ATP in dauer-defective *daf-2* and *daf-7* mutants, progressively with age. In contrast, we did not detect ATP elevation in prohibitin-depleted wild-type animals during ageing (Fig. 3a; data for day 10 shown). The significant energy surplus in *daf-2* mutants lacking prohibitin correlates with their exceptionally long lifespan (Fig. 1b). Our findings indicate that prohibitin moderates ATP levels under conditions of reduced diapause signalling.

What is the molecular basis of the different impact of prohibitin on ATP levels between wild type and diapause mutant animals? Mitochondrial energy metabolism is linked to fat utilization in both nematodes and mammals. We visualized fat depositions in the intestine of wild-type animals and diapause mutants during ageing, using the vital dye Nile red (see Methods). Fat accumulates during ageing in wild-type animals and in two representative diapause mutants (*daf-2* and *daf-7*; Fig. 3b and Supplementary Fig. 3a). These observations were confirmed by Sudan black staining of fixed animals (Supplementary Fig. 3b). Prohibitin deficiency markedly reduces intestinal fat content early in adulthood, in all genetic backgrounds (Fig. 3b and Supplementary Fig. 3a, day 5). However, the effect of prohibitin depletion diminishes with age in wild-type animals, whereas it remains strong in both diapause mutants (Fig. 3b and Supplementary Fig. 3a, day 10 and day 15). Thus, prohibitin differentially modulates animal fat content in a genetic background- and age-specific manner.

The nuclear hormone receptor NHR-49 is a key regulator of fat mobilization, modulating fat consumption and maintaining a normal balance of fatty acid saturation. Elimination of NHR-49 causes fat accumulation due to reduced expression of fatty acid β -oxidation enzymes such as the delta-9 stearyl-CoA desaturase FAT-7, which is required for the synthesis of monounsaturated fatty acids¹³. Prohibitin deficiency extends the lifespan of both *nhr-49* and *fat-7* mutants (Fig. 3c, d; Supplementary Table 1). In addition, knockdown of prohibitin reduces intestinal fat in *nhr-49* and *fat-7* mutant animals (Supplementary Fig. 4a, b). Taken together, our findings indicate that prohibitin deficiency engages fat metabolism to promote longevity.

Prohibitin has been implicated in several human cancers and is generally overexpressed in transformed cells compared with their non-transformed counterparts⁵. We examined the requirement for prohibitin activity during tumour formation in *C. elegans*. Although *C. elegans* somatic cells are post-mitotic, germ cells are continually dividing during oogenesis. *gld-1* is a tumour suppressor gene that encodes a protein containing a K homology RNA-binding domain that is required for meiotic cell cycle progression during oogenesis¹⁴. *gld-1* mutant animals develop lethal germline tumours and are short lived because of ectopic germ cell overproliferation in the gonad¹⁴. Germ cells eventually leak out of the gonad into the body cavity or, through the vulva, to the outside. We find that prohibitin deficiency blocks tumour formation and extends lifespan in *gld-1* mutants (compare Supplementary Fig. 5a with b; Supplementary Table 1). These observations indicate a critical function of prohibitin in actively proliferating cells^{6,15,16}.

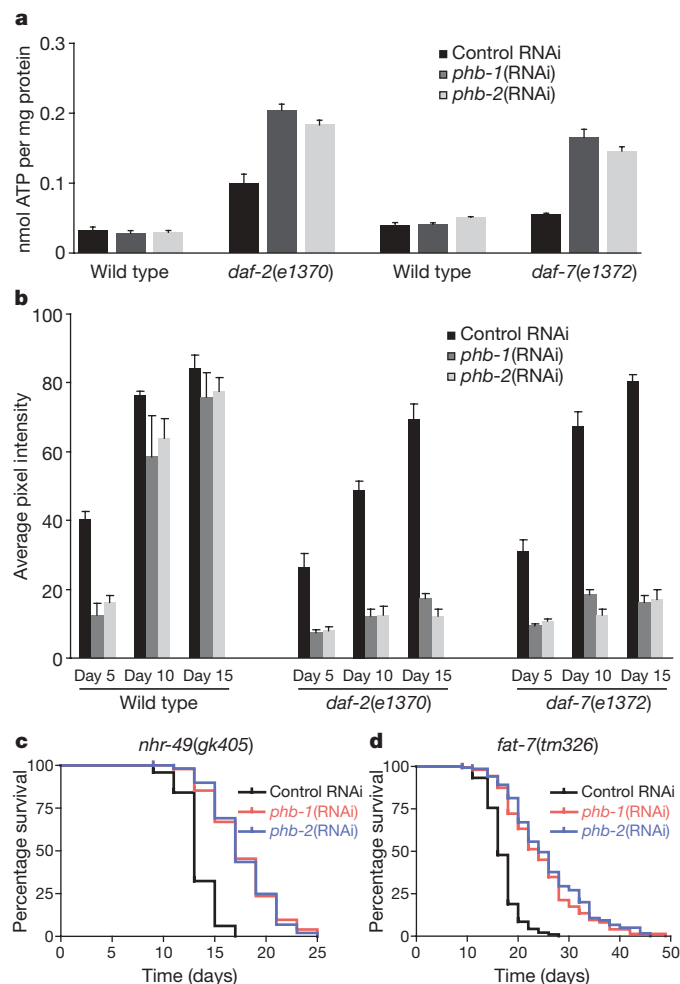


Figure 3 | Effects of prohibitin depletion on energy metabolism. **a**, RNAi knockdown of either *phb-1* or *phb-2* specifically increases ATP levels in dauer-defective *daf-2* or *daf-7* mutants during ageing (day 10 of adulthood). No effect is observed in wild-type (N2) animals (error bars denote standard deviation; $P < 0.005$, unpaired t test; assays were carried out at 20 °C). **b**, Quantification of intestinal fluorescence after Nile red staining of wild-type animals and dauer-defective mutants, subjected to RNAi with either *phb-1* or *phb-2* at day 5, day 10 and day 15 of adulthood (error bars denote standard deviation; $P < 0.005$, unpaired t test; assays were carried out at 20 °C; see Methods). **c**, **d**, Survival curves of short-lived *nhr-49* mutants (**c**) and animals lacking the delta-9 desaturase *FAT-7* (**d**), subjected to RNAi with either *phb-1* or *phb-2*.

To investigate the mechanism by which prohibitin influences mitochondrial activity to modulate longevity, we analysed cellular mitochondrial content during ageing. We found that prohibitin elimination promotes adult-onset mitochondrial proliferation in intestinal fat-storing cells of wild-type animals whereas, strikingly, it reduces mitochondrial content in diapause mutants. Notably, these mutants contain less mitochondria compared to wild type, during late adulthood (Fig. 4a and Supplementary Fig. 6a, b). Mitochondrial and fat content is also reduced upon knockdown of *phb* genes in mutants with compromised mitochondrial function, in dietary restricted animals, and in fat metabolism mutants (Fig. 4b and Supplementary Fig. 7a). In *Drosophila* and in the adipose tissue of mice, FOXO transcription factors inhibit mitochondrial proliferation^{17,18}. We observed a similar effect in *C. elegans* diapause mutants (Fig. 4a), where DAF-16/FOXO is derepressed⁹. In contrast, prohibitin depletion does not significantly alter fat or mitochondrial content in animals lacking DAF-16 ($P > 0.1$, unpaired t test; Supplementary Fig. 8a). Hence, DAF-16 is required to mediate the effects of prohibitin deficiency on fat and mitochondrial content in wild-type animals. However, elimination of prohibitin

during adulthood does not affect the subcellular localization of either a wild-type or a constitutively nuclear DAF-16 reporter fusion (Supplementary Fig. 8b). In addition, constitutive nuclear localization of DAF-16 is not sufficient to extend the lifespan of prohibitin-depleted animals (Supplementary Table 1). Moreover, the ageing effects of knockout or overexpression of sirtuin SIR-2.1, a regulator of DAF-16, are independent of prohibitin (Supplementary Table 1).

Together, our findings show a new mechanism that couples nutrient availability and diapause signals with energy metabolism during ageing. We hypothesize that prohibitin normally functions to promote longevity by moderating fat utilization and energy production via the mitochondrial respiratory chain. Under conditions that favour diapause, such as limited nutrient availability, cells adapt by shifting towards fermentative metabolism¹⁹. Under such conditions, when energy demands exceed the capacity of mitochondrial respiration, prohibitin deficiency is beneficial for survival. We tested this hypothesis by monitoring survival of animals lacking prohibitin at a higher temperature (25 °C), where metabolic activity is elevated and energy demand is higher. Notably, whereas knockdown of prohibitin shortens lifespan at 20 °C it extends lifespan at 25 °C (Supplementary Table 1). We also tested the requirement for prohibitin under acute thermal stress (35 °C). Prohibitin deficiency renders wild-type animals strongly thermotolerant and further enhances the thermotolerance of *daf-2* diapause mutants (Supplementary Fig. 9a, b). Thus, the metabolic state determines whether prohibitin will promote or compromise longevity. Increased thermotolerance is independent of DAF-16/FOXO (Supplementary Fig. 8c), indicating that other pathways are involved in metabolic changes elicited by prohibitin deficiency.

We also examined the effects of prohibitin depletion on animals under oxidative stress. We induced oxidative stress by using sodium azide (NaN_3), a potent and specific inhibitor of cytochrome *c* oxidase, a component of the mitochondrial electron transport chain complex IV. Prohibitin depletion enhances survival after treatment with sodium azide during adulthood in diapause mutants (Supplementary Fig. 9c). Similarly, prohibitin deficiency increases resistance of adult *daf-2* mutant animals to the herbicide paraquat (N,N' -dimethyl-4,4'-bipyridinium dichloride), a generator of superoxide anions (Supplementary Fig. 9d). By contrast, lack of prohibitin compromises paraquat resistance during adulthood in an otherwise wild-type genetic background and during L4 larval development in both wild-type and *daf-2* mutant animals (Supplementary Fig. 9e, f; ref. 6). Therefore, elimination of prohibitin under conditions of reduced diapause signalling further increases oxidative stress resistance during adulthood. We conclude that lack of prohibitin diminishes mitochondrial proliferation during ageing, augments oxidative stress resistance and extends lifespan, specifically under conditions of reduced insulin/IGF and TGF- β signalling.

Mitochondria are the main sites of reactive oxygen species generation within cells. We measured reactive oxygen species formation upon prohibitin depletion in both wild-type and *daf-2* mutant adult animals under normal and oxidative stress conditions. Reactive oxygen species levels were slightly lower in *daf-2* mutants compared to wild type (Supplementary Fig. 9g). Notably, although prohibitin deficiency increased reactive oxygen species formation in wild-type animals, it reduced reactive oxygen species levels in *daf-2* mutant adults (Supplementary Fig. 9g). We also assessed mitochondrial membrane potential and oxygen consumption in prohibitin-deficient, wild-type and *daf-2* mutant animals. Knockdown of prohibitin slightly reduces mitochondrial membrane potential, while selectively increasing oxygen consumption in *daf-2* mutants (Supplementary Fig. 10a, b). Paradoxically, reactive oxygen species formation has been shown to underlie oxidative stress resistance and lifespan extension under glucose restriction in *C. elegans*²⁰. It has been suggested that stress resistance and longevity are due to induction of a hormetic response (mitohormesis)²⁰. We investigated whether increased reactive oxygen species formation augments stress resistance and extends lifespan in prohibitin-deficient animals under stress. We treated prohibitin-depleted

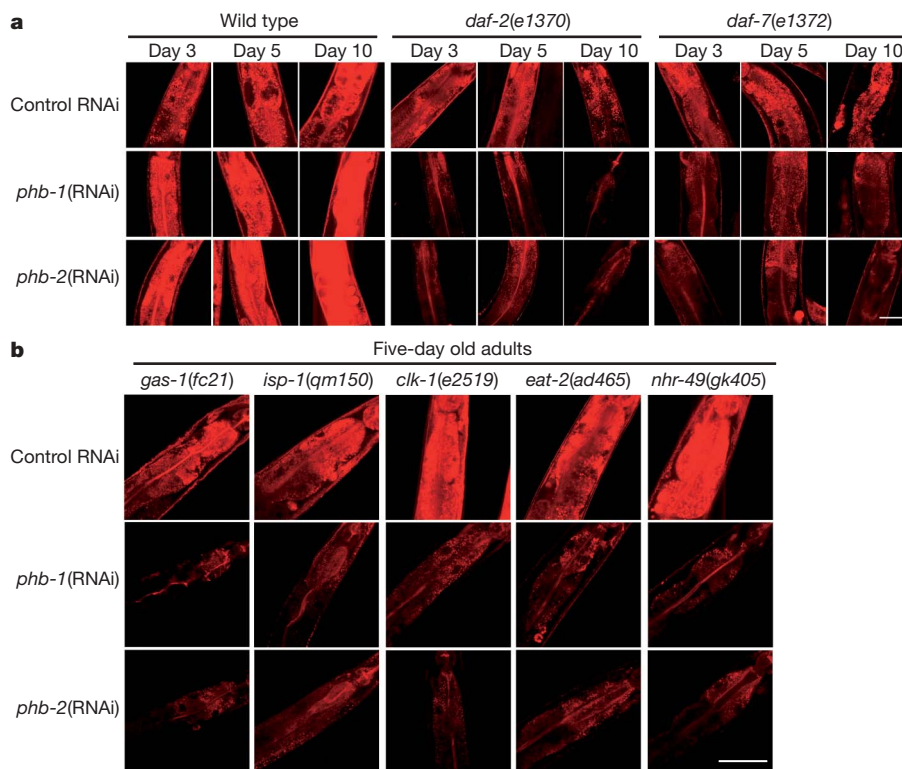


Figure 4 | Prohibitin depletion and intestinal fat-storing cell mitochondrial content. **a**, MitoTracker Deep Red 633 staining of intestinal mitochondria in wild-type animals and dauer-defective *daf-2* or *daf-7* mutants subjected to RNAi with either *phb-1* or *phb-2*. Images were acquired under the same exposure, using a $\times 40$ objective lens, at day 3, day 5 and day 10 of adulthood (see Methods; the anterior part of the intestine is shown, the head is located

at the top; bar, 50 μ m). **b**, MitoTracker Deep Red 633 staining of intestinal mitochondria in *gas-1(fc21)*, *isp-1(qm150)*, *clk-1(e2519)*, *eat-2(ad465)* and *nhr-49(gk405)* mutants subjected to RNAi with either *phb-1* or *phb-2*. Images were acquired under the same exposure, using a $\times 40$ objective lens at day 5 of adulthood.

animals experiencing either oxidative stress (*mev-1* mutants) or thermal stress (wild-type nematodes grown at 25 °C) with *N*-acetylcysteine, a compound that functions as a free-radical scavenger²⁰. We found no effect on longevity conferred by prohibitin knockdown (Supplementary Table 1 and Supplementary Fig. 11). Thus mitohormesis is unlikely to mediate the effects of prohibitin elimination on ageing.

The AMP-dependent kinase (AMPK) AAK-2 has been implicated in coupling energy levels and insulin/IGF-1 signals to modulate lifespan in *C. elegans*^{20,21}. AMPK targets p53 to promote cell survival under conditions of nutrient deprivation²². We find that prohibitin deficiency shortens the lifespan of both *aak-2* and p53 (*cep-1*) mutant animals (Supplementary Table 1 and Supplementary Fig. 12a, b). Interestingly, *aak-2* mutants contain more mitochondria than wild-type animals. Depletion of prohibitins further increases mitochondrial proliferation in this genetic background (Supplementary Fig. 12c, d). In addition, AAK-2 deficiency ameliorates fat content reduction upon prohibitin depletion (Supplementary Fig. 12e, f), indicating that AAK-2 is involved in mediating prohibitin effects on fat content. We also investigated the involvement of the mitogen-activated protein kinase (MAPK) JNK-1, which promotes DAF-16/FOXO nuclear localization under conditions of stress²³, and the Akt/PKB homologue AKT-1, which transduces insulin/IGF-1 signals²⁴, in mediating the effects of prohibitin depletion on metabolism and ageing. We found that *phb* gene knockdown marginally extends the lifespan of animals overexpressing *jnk-1*, whereas it shortens lifespan in animals lacking JNK-1 (Supplementary Table 1 and Supplementary Fig. 13a). Overexpression of JNK-1 in wild-type animals reduces mitochondrial content and suppresses mitochondrial proliferation upon prohibitin depletion. By contrast, mitochondrial content is higher in *jnk-1* mutants and remains unchanged after prohibitin removal (Supplementary Fig. 13b, c). Similarly, prohibitin deficiency does not alter fat content in *jnk-1* mutant animals (Supplementary

Fig. 13d, e). Thus, JNK-1 is required for mitochondrial proliferation and reduction of fat content in animals lacking prohibitin. Elimination of prohibitin does not shorten the lifespan of animals without AKT-1 (Supplementary Table 1 and Supplementary Fig. 13a). Mitochondrial content is reduced in *akt-1* mutants and does not increase upon prohibitin depletion, compared to wild type (Supplementary Fig. 13b, c). In contrast, fat content in *akt-1* mutants is higher than in wild type and sharply diminishes in the absence of prohibitin, similarly to *daf-2* mutant animals (Supplementary Fig. 13d, e). Taken together, our observations indicate that the JNK-1 kinase, in part, mediates the effects of prohibitin deficiency on fat metabolism and mitochondrial proliferation. This response is potentiated under conditions of low diapause signalling, where AKT-1 activity is reduced and DAF-16/FOXO nuclear localization is not blocked.

In mammalian cells and in *C. elegans*, loss of prohibitin disrupts the reticular mitochondrial network and leads to accumulation of fragmented mitochondria^{6,25}. Prohibitin maintains mitochondrial integrity and biogenesis by stabilizing the dynamin-like GTPase OPA1²⁵, a core component of the mitochondrial fusion machinery, which is required for mitochondrial fusion and cristae maintenance, and has been implicated in the pathogenesis of inherited autosomal dominant optic atrophy²⁶. The *eat-3* gene encodes the *C. elegans* homologue of OPA1. We examined whether prohibitin functions through OPA1 to regulate metabolism during ageing, by analysing the effects of EAT-3 depletion on fat metabolism, mitochondrial content, ATP levels and lifespan, in wild type and insulin/IGF1 signalling-deficient animals. In contrast to prohibitin depletion, loss of EAT-3 extends the lifespan of otherwise wild-type animals, whereas it shortens the lifespan of long-lived *daf-2* mutants (Supplementary Table 1 and Supplementary Fig. 14a, b). We did not observe significant alterations of fat metabolism, ATP levels or mitochondrial content in

EAT-3-deficient animals (Supplementary Fig. 14c–e). Our findings indicate that prohibitin functions independently of EAT-3/OPA1 to regulate metabolism during ageing.

What is the origin of prohibitin effects on metabolism and ageing? We propose that under normal conditions the mitochondrial prohibitin complex promotes longevity by moderating fat metabolism, mitochondrial proliferation and energy levels in *C. elegans*. In addition to maintaining normal mitochondrial metabolism, prohibitin functions as a negative regulator of mitochondrial proliferation in wild-type animals (Supplementary Fig. 15). Elimination of prohibitin may activate a cellular retrograde response that induces mitochondrial overproliferation. In turn, accumulation of defective mitochondria lacking prohibitin results in increased reactive oxygen species production, metabolic defects, consequent cellular damage and reduced lifespan. Interestingly, prohibitin depletion elicits reactive-oxygen-species-dependent, Akt hyperactivation in endothelial cells²⁷. Under low diapause signalling and stress conditions, where AKT-1-mediated inhibition of DAF-16/FOXO nuclear localization is relieved, the JNK-1 and AAK-2 stress-related signalling pathways are activated by prohibitin depletion, adjusting cellular metabolism towards fat utilization and promoting longevity.

Our study reveals an important role of prohibitin in regulating fat metabolism and mitochondrial proliferation during ageing. The opposing effects of prohibitin on longevity indicate that specific cellular mechanisms may differentially regulate ageing, depending on extrinsic or intrinsic cues such as diapause signalling or energy demands. The tight evolutionary conservation and ubiquitous expression of prohibitin proteins indicate a similar role during ageing in other organisms.

METHODS SUMMARY

Lifespan analysis. Lifespan assays were performed at 20 °C unless noted otherwise. Synchronous animal populations were generated by hypochlorite treatment of gravid adults to obtain tightly synchronized embryos that were allowed to develop into adulthood under appropriate, defined conditions. Animals were transferred to fresh plates in groups of 10–20 worms per plate for a total of 100–150 individuals per experiment. The day of egg harvest was used as $t = 0$. Animals were transferred to fresh plates every 2–4 days thereafter and were examined every day for touch-provoked movement and pharyngeal pumping, until death. Survival curves were generated using the product-limit method of Kaplan and Meier. The log-rank (Mantel–Cox) test was used to evaluate differences in survival and determine P values.

MitoTracker staining. Animals were stained overnight on plates containing MitoTracker Deep Red 633 at a final concentration of 100 nM. Animals were mounted on 2% agarose pads in M9 buffer containing 10 mM sodium azide and scanned at room temperature with a 637 nm laser beam, under a confocal microscope.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 10 July; accepted 24 August 2009.

1. Roberts, S. B. & Rosenberg, I. Nutrition and aging: changes in the regulation of energy metabolism with aging. *Physiol. Rev.* **86**, 651–667 (2006).
2. Balaban, R. S., Nemoto, S. & Finkel, T. Mitochondria, oxidants, and aging. *Cell* **120**, 483–495 (2005).
3. Back, J. W. *et al.* A structure for the yeast prohibitin complex: structure prediction and evidence from chemical crosslinking and mass spectrometry. *Protein Sci.* **11**, 2471–2478 (2002).
4. Nijtmans, L. G. *et al.* Prohibitins act as a membrane-bound chaperone for the stabilization of mitochondrial proteins. *EMBO J.* **19**, 2444–2451 (2000).
5. Mishra, S., Murphy, L. C., Nyomba, B. L. & Murphy, L. J. Prohibitin: a potential target for new therapeutics. *Trends Mol. Med.* **11**, 192–197 (2005).
6. Artal-Sanz, M. *et al.* The mitochondrial prohibitin complex is essential for embryonic viability and germline function in *Caenorhabditis elegans*. *J. Biol. Chem.* **278**, 32091–32099 (2003).

7. Kimura, K. D., Tissenbaum, H. A., Liu, Y. & Ruvkun, G. *daf-2*, an insulin receptor-like gene that regulates longevity and diapause in *Caenorhabditis elegans*. *Science* **277**, 942–946 (1997).
8. Kenyon, C. The plasticity of aging: insights from long-lived mutants. *Cell* **120**, 449–460 (2005).
9. Shaw, W. M. *et al.* The *C. elegans* TGF- β dauer pathway regulates longevity via insulin signaling. *Curr. Biol.* **17**, 1635–1645 (2007).
10. Estevez, M. *et al.* The *daf-4* gene encodes a bone morphogenetic protein receptor controlling *C. elegans* dauer larva development. *Nature* **365**, 644–649 (1993).
11. Ren, P. *et al.* Control of *C. elegans* larval development by neuronal expression of a TGF- β homolog. *Science* **274**, 1389–1391 (1996).
12. Li, W., Kennedy, S. G. & Ruvkun, G. *daf-28* encodes a *C. elegans* insulin superfamily member that is regulated by environmental cues and acts in the DAF-2 signaling pathway. *Genes Dev.* **17**, 844–858 (2003).
13. Van Gilst, M. R., Hadjivassiliou, H., Jolly, A. & Yamamoto, K. R. Nuclear hormone receptor NHR-49 controls fat consumption and fatty acid composition in *C. elegans*. *PLoS Biol.* **3**, e53 (2005).
14. Francis, R., Barton, M. K., Kimble, J. & Schedl, T. *gld-1*, a tumor suppressor gene required for oocyte development in *Caenorhabditis elegans*. *Genetics* **139**, 579–606 (1995).
15. Coates, P. J., Jamieson, D. J., Smart, K., Prescott, A. R. & Hall, P. A. The prohibitin family of mitochondrial proteins regulate replicative lifespan. *Curr. Biol.* **7**, 607–610 (1997).
16. Coates, P. J. *et al.* Mammalian prohibitin proteins respond to mitochondrial stress and decrease during cellular senescence. *Exp. Cell Res.* **265**, 262–273 (2001).
17. Gershman, B. *et al.* High-resolution dynamics of the transcriptional response to nutrition in *Drosophila*: a key role for dFOXO. *Physiol. Genom.* **29**, 24–34 (2007).
18. Nakae, J. *et al.* Forkhead transcription factor FoxO1 in adipose tissues regulates energy storage and expenditure. *Diabetes* **57**, 563–576 (2008).
19. Rea, S. & Johnson, T. E. A metabolic model for life span determination in *Caenorhabditis elegans*. *Dev. Cell* **5**, 197–203 (2003).
20. Schulz, T. J. *et al.* Glucose restriction extends *Caenorhabditis elegans* life span by inducing mitochondrial respiration and increasing oxidative stress. *Cell Metab.* **6**, 280–293 (2007).
21. Apfeld, J., O'Connor, G., McDonagh, T., DiStefano, P. S. & Curtis, R. The AMP-activated protein kinase AAK-2 links energy levels and insulin-like signals to lifespan in *C. elegans*. *Genes Dev.* **18**, 3004–3009 (2004).
22. Jones, R. G. *et al.* AMP-activated protein kinase induces a p53-dependent metabolic checkpoint. *Mol. Cell* **18**, 283–293 (2005).
23. Oh, S. W. *et al.* JNK regulates lifespan in *Caenorhabditis elegans* by modulating nuclear translocation of forkhead transcription factor/DAF-16. *Proc. Natl Acad. Sci. USA* **102**, 4494–4499 (2005).
24. Paradis, S. & Ruvkun, G. *Caenorhabditis elegans* Akt/PKB transduces insulin receptor-like signals from AGE-1 PI3 kinase to the DAF-16 transcription factor. *Genes Dev.* **12**, 2488–2498 (1998).
25. Merkwirth, C. *et al.* Prohibitins control cell proliferation and apoptosis by regulating OPA1-dependent cristae morphogenesis in mitochondria. *Genes Dev.* **22**, 476–488 (2008).
26. Olichon, A. *et al.* Loss of OPA1 perturbs the mitochondrial inner membrane structure and integrity, leading to cytochrome c release and apoptosis. *J. Biol. Chem.* **278**, 7743–7746 (2003).
27. Schleicher, M. *et al.* Prohibitin-1 maintains the angiogenic capacity of endothelial cells by regulating mitochondrial function and senescence. *J. Cell Biol.* **180**, 101–112 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Pasparki for technical support with experiments. Some nematode strains used in this work were provided by the *C. elegans* Gene Knockout Project at OMRF (<http://www.mutantfactory.ouhsc.edu/>), which is part of the International *C. elegans* Gene Knockout Consortium, the *Caenorhabditis* Genetics Center, which is funded by the NIH National Center for Research Resources (NCRR), and S. Mitani (National Bioresource Project) in Japan. We thank A. Fire for plasmid vectors and J. Berden for antibodies. This work was funded by grants from EMBO, the European Research Council (ERC), the Marie Curie Fellowships Programme and the European Commission Coordination Action ENINET (contract number LSHM-CT-2005-19063).

Author Contributions M.A.-S. and N.T. designed and performed experiments, analysed data and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to N.T. (tavernarakis@imbb.forth.gr) or M.A.-S. (martal@ibv.csic.es).

METHODS

Strains and genetics. We followed standard procedures for *C. elegans* strain maintenance. Nematode rearing temperature was kept at 20 °C, unless noted otherwise. The following strains were used in this study: N2, wild-type Bristol isolate; CB1370, *daf-2(e1370)III*; CB1372, *daf-7(e1372)III*; CB4876: *clk-1(e2519)III*; CF1139: *daf-16(mu86)I;muIs61*; CF1308: *daf-16(mu86)I;muEx116*; CF1371: *daf-16(mu86)I;muEx151*; CW152: *gas-1(fc21)X*; DA465: *eat-2(ad465)II*; DR26: *daf-16(m26)I*; DR47: *daf-11(m47)V*; DR63: *daf-4(m63)III*; DR1309: *daf-16(m26)I;daf-2(e1370)II*; GR1307: *daf-16(mgDf50)I*; HT941: *lpIn1*; JK1466: *gld-1(q485)/dpy-5(e61)unc-13(e51)I*; LG100: *geln3*; MQ887: *isp-1(qm150)IV*; RB754: *aak-2(ok524)X*; RB759: *akt-1(ok525)V*; TK22: *mev-1(kn1)III*; VC8: *jnk-1(gk7)IV*; VC199: *sir-2.1(ok434)IV*; VC870: *nhr-49(gk405)I;fat-7(tm0326)*; XY1054: *cep-1(lg12501)I*; N2Ex[p_{phb-1}-PHB-1::GFP pRF4], N2Ex[p_{phb-2}-PHB-2::GFP pRF4] and *phb-1(tm2571)I;sDp2(lf)*. The *sDp2(lf)* balancer chromosomal duplication is unstable during meiosis and is lost in about 30% of the progeny, which become sterile homozygous *phb-1(tm2571)* adults.

Molecular cloning. For engineering *phb-1* and *phb-2* dsRNA-producing *Escherichia coli* bacteria, the corresponding genomic DNA fragments, previously inserted into the pBluescript II plasmid vector⁶, were excised by *SacI*/*KpnI* and *KpnI*/*SpeI*, respectively, and sub-cloned into the pL4440 RNAi vector. The resulting plasmid construct was used to transform HT115(DE3) *E. coli* bacteria, deficient for RNase-E. Bacteria carrying an empty vector were used in control experiments. To generate p_{phb-1}-PHB-1::GFP and p_{phb-2}-PHB-2::GFP full-length GFP reporter fusions, DNA fragments derived from the *phb-1* and *phb-2* loci were PCR-amplified using appropriate oligonucleotide primers and fused to GFP. For p_{phb-1}-PHB-1::GFP, the primers 5'-AACTGCAGCTCAACGCGTGAGCCATACC-3' and 5'-GCTCTAGAGGATTGAGGTTGAGAGG-3' were used to amplify a 1.7 kb DNA fragment encompassing the promoter plus the full coding region of *phb-1*, which was digested with *PstI*/*XbaI* and inserted into plasmid vector pPD95.77. Similarly, for p_{phb-2}-PHB-2::GFP, the primers 5'-ACATGCATCGCAGTCAGAGATAAAGACCG-3' and 5'-GCTCTAGAGCGTCTTTGTCGGTACAC-3' were used to amplify a 1.9 kb DNA fragment encompassing the promoter plus the full coding region of *phb-2*, which was digested with *SphI*/*XbaI* and inserted into pPD95.77. Reporter constructs were injected into the gonads of wild-type animals together with pRF4, a plasmid that carries the *rol-6(su1006)* dominant transformation marker. Two independent, transgenic lines were obtained for the p_{phb-1}-PHB-1::GFP plasmid construct and roller hermaphrodites were examined for reporter fusion expression. For p_{phb-2}-PHB-2::GFP, we obtained and examined numerous (>50) F1 transgenic progeny, none of which propagated to generate a stable transgenic line, probably because the GFP moiety of the fusion interferes with the formation of the complex²⁸. Transgenic animals were mounted on a 2% agarose pad in M9 buffer, containing 10 mM sodium azide and scanned at room temperature with a 488 nm laser beam, under a confocal microscope (Zeiss AxioScope with a Bio-Rad Radiance 2100 scanhead). Images were acquired using a 515 ± 15 nm band-pass filter and a ×40 Plan-NEOFLUAR objective (numerical aperture 0.75).

Lifespan analysis. Lifespan assays were performed at 20 °C unless noted otherwise. Synchronous animal populations were generated by hypochlorite treatment of gravid adults to obtain tightly synchronized embryos that were allowed to develop into adulthood under appropriate, defined conditions. For RNAi lifespan experiments worms were placed on NGM plates containing 1–2 mM IPTG and seeded with HT115(DE3) bacteria transformed with either the pL4440 vector or the test RNAi construct. Progeny were grown at 20 °C unless noted otherwise, through the L4 larval stage and then transferred to fresh plates in groups of 10–20 worms per plate for a total of 100–150 individuals per experiment. The day of egg harvest and initiation of RNAi was used as *t* = 0. Animals were transferred to fresh plates every 2–4 days thereafter and were examined every day for touch-provoked movement and pharyngeal pumping, until death. Worms that died due to internally hatched eggs, an extruded gonad or desiccation due to crawling on the edge of the plates, were censored and incorporated as such into the data set. Each survival assay was repeated at least three times and figures represent typical assays. Survival curves were created using the product-limit method of Kaplan and Meier. The log-rank (Mantel–Cox) test was used to evaluate differences between survivals and determine *P* values. We used the Prism software package (GraphPad Software) to carry out statistical analysis and to determine lifespan values.

Stress resistance assays. To evaluate thermotolerance, four-day-old adult hermaphrodites were placed on pre-warmed (35 °C) NGM plates and incubated at 35 °C. At the indicated times, plates were removed and worms were scored for motility, provoked movement and pharyngeal pumping. Worms failing to display any of these traits were scored as dead. Three distinct populations of 30 adults were scored repeatedly over the assay period. Statistical tests were performed using the Kaplan–Meier survival analysis, as described above for lifespan data. To analyse oxidative stress resistance, 7-day-old adults were exposed to 1 mM

sodium azide (Sigma-Aldrich) for 18 h, on RNAi plates, at 20 °C. Animals were scored for survival after a 3-h recovery period. To assay paraquat resistance, 7-day-old adults were exposed to 40 mM paraquat (Aldrich) on RNAi plates at 20 °C and survival was scored from day 8 of adulthood. For paraquat resistance of L4 larvae, animals were exposed to 2 mM paraquat on RNAi plates at 20 °C and survival was scored every 2 days. The percentage of surviving animals for each drug treatment was calculated in three independent experiments. In each experiment, 100 animals were analysed. Statistical analysis of data was performed using the Excel software package (Microsoft).

Fat staining. Nile red powder (catalogue number N3013, Sigma-Aldrich) was dissolved in DMSO at 5 mg ml⁻¹, diluted in M9 and added on top of nematode growth media (NGM) plates seeded with HT115(DE3) *E. coli* bacteria harbouring the appropriate RNAi plasmids, to a final concentration of 0.02 µg ml⁻¹. Synchronous embryos were allowed to develop into adulthood and grow continuously on Nile red-containing plates. The extent of fat staining was assessed at specific time points by epifluorescence microscopy²⁹. Animals were observed using a ×20 Plan-NEOFLUAR objective (numerical aperture 0.50), coupled with a 546 ± 12 nm band-pass excitation and a 590 nm long-pass emission filter, on a Zeiss AxioPlan microscope (Carl Zeiss). Images were acquired using a Zeiss AxioCam digital colour camera. Emission intensity was measured on greyscale images with a pixel depth of 8 bit (256 shades of grey). Average pixel intensity values were calculated by sampling three images of different animals, three times each (nine measurements total for each strain/condition). We calculated the mean and maximum pixel intensity for each animal in these images using the ImageJ software (<http://rsb.info.nih.gov/ij/>). For each experiment, at least 50 images were processed over at least five independent trials. Because recent studies suggested that Nile red may not accurately indicate fat content in insulin/IGF-1 mutants, owing to uptake and/or anatomical issues³⁰, animals were also stained with Sudan black (Sigma-Aldrich) as described previously⁷. Briefly, non-starved animals were collected in M9 buffer and washed three times. Animals were then fixed by adding 10% paraformaldehyde solution to final concentration of 1%. Fixed animals were frozen at -80 °C and underwent three cycles of freeze–thawing before washing in cold M9 buffer three times. Animals were subsequently dehydrated by ethanol washes (serially in 25%, 50% and 70% ethanol). For staining, three volumes of saturated Sudan black B solution (in 70% ethanol) were added to worms. Animals were incubated overnight and washed thrice with 70% ethanol before observation. Animals were observed using a ×20 Plan-NEOFLUAR objective (numerical aperture 0.50) on a Zeiss AxioPlan microscope (Carl Zeiss). Images were acquired using a Zeiss AxioCam digital camera. Emission intensity was measured on greyscale images with a pixel depth of 8 bit (256 shades of grey). Average pixel intensity values were calculated by sampling three images of different animals, three times each (nine measurements total for each strain). We calculated the mean and maximum pixel intensity for each animal in these images using the ImageJ software (<http://rsb.info.nih.gov/ij/>). Numbers obtained were subtracted from 255 to obtain the values depicted in Supplementary Fig. 3. For each experiment, at least 20 images were processed over at least three independent trials.

MitoTracker staining. Animals were stained overnight on RNAi plates containing MitoTracker Deep Red 633 (catalogue number M-22426; Molecular Probes, Invitrogen) at a final concentration of 100 nM. Animals were mounted in a 2% agarose pad in M9 buffer containing 10 mM sodium azide and scanned at room temperature with a 637 nm laser beam, under a confocal microscope (Zeiss AxioPlan coupled to a Bio-Rad Radiance 2000 laser scanning system). Images of emission were acquired using a 660 nm long-pass filter and a ×40 Plan-NEOFLUAR objective (NA 0.75), and processed with Bio-Rad LaserSharp 2000 software.

ATP measurements. To determine ATP content, 50 age-matched animals were collected in 50 µl of S Basal buffer and frozen at -80 °C. Nematodes were collected at the L4 stage of development and on day 2, day 10 and day 15 of adulthood. Frozen worms were immersed in boiling water for 15 min, cooled and centrifuged to pellet insoluble debris. The supernatant was moved to a fresh tube and diluted tenfold before measurement. ATP content was determined by using the Roche ATP bioluminescent assay kit HSII (Roche Applied Science) and a TD-20/20 luminometer (Turner Designs). ATP levels were normalized to total protein content.

Quantification of reactive oxygen species production. Reactive oxygen species formation was quantified as described^{20,31}. Briefly, we use the membrane-permeable non-fluorescent dye 2,7-dichlorodihydrofluorescein-diacetate (H2-DCF-DA) (Sigma-Aldrich). H2-DCF-DA is deacetylated and becomes membrane impermeable after entering the cell. H2-DCF fluoresces upon oxidation to 2,7-dichlorofluorescein (DCF) by reactive oxygen species. Young adults treated as described above were washed off of the plates with M9 buffer. After washing to reduce bacterial content, a 50 µl volume of worm suspension was pipetted in four replicates into the wells of a 96-well plate with opaque walls and bottom and

allowed to equilibrate to room temperature. A fresh 100 μ M H₂-DCF-DA solution (50 μ l) was pipetted to the suspensions, resulting in a final concentration of 50 μ M. Basal fluorescence was measured after addition of H₂-DCF-DA, in a microplate reader at excitation/emission wavelengths of 485 and 520 nm. Plates were kept for 1 h shaking at 20 °C. Then, a second measurement was performed. The initial fluorescence and the fluorescence signals of control wells were subtracted from the second measurement. Values were normalized to protein content.

Electrophoresis and western blot analysis. For one-dimensional SDS–PAGE, worm pellets were re-suspended in five volumes of SDS-sample buffer, boiled for 5 min, and the proteins were resolved on 15% gels. Following electrophoresis, proteins were blotted to PVDF membranes, and immunoreactive material was visualized by chemiluminescent detection (ECL; Amersham) according to the manufacturer's instructions. A polyclonal antibody raised against the 25 carboxy-terminal amino acids of the murine PHB-1 protein has been described previously¹⁵. Polyclonal antibody against the yeast β -subunit of F₁-ATPase was a gift from J. Berden. Anti-actin antibody was obtained from ICN (clone C4) and used at a dilution of 1:10,000.

Mitochondrial DNA quantification. Mitochondrial DNA (mtDNA) was quantified using quantitative real time PCR as described previously³². We used the primers 5'-GTTTATGCTGCTGCTAGCGTG-3' and 5'-CTGTTAAAGCAAGTG GACGAG-3' (Mito1 set) for mtDNA. The results were normalized to genomic DNA using the following primers specific for *ama-1*: 5'-TGGAACCTCTGGA GTCACACC-3' and 5'-CATCCTCCTTCATTGAACGG-3'. Quantitative PCR was performed using the Bio-Rad CFX96 Real-Time PCR system, and was repeated three times.

Oxygen consumption rate measurements. Oxygen consumption rates were measured as previously described³³ using a Clark-type electrode with some minor modifications (Hansatech Instruments). Young adult worms were washed and collected in S-basal buffer. Approximately 100 μ l of slurry pellet of worms were delivered into the chamber in 3 ml of S-basal medium. The chamber was kept at 25 °C, and measurements were done for 5–15 min, depending on the oxygen consumption rate. The slope of the straight portion of the plot was used to derive the oxygen consumption rate. Worms were recovered after respiration measurements and collected for protein quantification. Rates were normalized to protein content. We performed three independent measurements per strain. Statistic analysis was performed using the Excel software package (Microsoft).

Membrane potential measurements. Mitochondrial membrane potential was measured *in vivo* using the fluorescent, lipophilic carbocyanine dye, 3,3'-dipropylthiadicarbocyanine iodide (diS-C₃(3); Sigma-Aldrich), as described³⁴. Stained and washed worms were immobilized with Levamisole before mounting on 2% agarose pads for microscopic examination with a Zeiss AxioPlan microscope (Carl Zeiss) equipped with a Zeiss AxioCam digital colour camera. Images

were acquired under the same exposure. Average pixel intensity values were calculated by sampling three images of different animals, three times each (nine measurements total for each strain/condition). We calculated the mean and maximum pixel intensity for each animal in these images using the ImageJ software (<http://rsb.info.nih.gov/ij/>). For each experiment, at least 50 images were processed over at least five independent trials.

Prohibitin overexpression. *phb-1* and *phb-2* overexpression plasmids were constructed by PCR amplification of the *phb-1* and *phb-2* loci, using primers 5'-CTCAACGCGTGAGCCATACC-3' and 5'-CGACATCGGGGAATTGATTC-3' for *phb-1*, and primers 5'-CGAGTCAGAGATAAGACCG-3' and 5'-AACCG GGAATTACATTCCAG-3' for *phb-2*. The resulting 2.1 kb and 2.5 kb fragments for *phb-1* and *phb-2*, respectively, were inserted into the plasmid vector pCRII-TOPO (Invitrogen). The two constructs, either each alone or both, were injected into the gonads of wild-type animals, together with pPD118.33, a plasmid that carries a p_{myo-2}GFP reporter fusion as transformation marker (pharyngeal muscle GFP expression). pRF4, a plasmid that carries the *rol-6* (*su1006*) dominant transformation marker was also used. We have not been able to establish stable transgenic lines overexpressing either each or both prohibitin genes. Although F₁ transgenic animals expressing the co-injection marker were obtained, none of the F₁ transgenic progeny segregated *phb*-overexpressing F₂ transgenic animals, indicating that overexpression is causing lethality. Growing F₁ transgenic animals on *phb* RNAi plates (and thus quenching *phb* gene overexpression) allowed generation of F₂ transgenics. These animals stopped propagating once shifted onto regular OP50 plates.

28. Tatsuta, T., Model, K. & Langer, T. Formation of membrane-bound ring complexes by prohibitins in mitochondria. *Mol. Biol. Cell* **16**, 248–259 (2005).
29. Ashrafi, K. *et al.* Genome-wide RNAi analysis of *Caenorhabditis elegans* fat regulatory genes. *Nature* **421**, 268–272 (2003).
30. Soukas, A. A., Kane, E. A., Carr, C. E., Melo, J. A. & Ruvkun, G. Rictor/TORC2 regulates fat metabolism, feeding, growth, and life span in *Caenorhabditis elegans*. *Genes Dev.* **23**, 496–511 (2009).
31. Kampkötter, A. *et al.* Effects of the flavonoids kaempferol and fisetin on thermotolerance, oxidative stress and FoxO transcription factor DAF-16 in the model organism *Caenorhabditis elegans*. *Arch. Toxicol.* **81**, 849–858 (2007).
32. Cristina, D., Cary, M., Lunceford, A., Clarke, C. & Kenyon, C. A regulated response to impaired respiration slows behavioral rates and increases lifespan in *Caenorhabditis elegans*. *PLoS Genet.* **5**, e1000450 (2009).
33. Braeckman, B. P., Houthoofd, K., De Vreese, A. & Vanfleteren, J. R. Assaying metabolic activity in ageing *Caenorhabditis elegans*. *Mech. Ageing Dev.* **123**, 105–119 (2002).
34. Gášková, D., DeCorby, A. & Lemire, B. D. DiS-C₃(3) monitoring of *in vivo* mitochondrial membrane potential in *C. elegans*. *Biochem. Biophys. Res. Commun.* **354**, 814–819 (2007).

LETTERS

Genetic variation in *IL28B* and spontaneous clearance of hepatitis C virus

David L. Thomas^{1*}, Chloe L. Thio^{1*}, Maureen P. Martin^{2*}, Ying Qi², Dongliang Ge³, Colm O'hUigin², Judith Kidd⁴, Kenneth Kidd⁴, Salim I. Khakoo⁵, Graeme Alexander⁶, James J. Goedert⁷, Gregory D. Kirk⁸, Sharyne M. Donfield⁹, Hugo R. Rosen¹⁰, Leslie H. Tobler¹¹, Michael P. Busch¹¹, John G. McHutchison¹², David B. Goldstein³ & Mary Carrington^{2,13}

Hepatitis C virus (HCV) infection is the most common blood-borne infection in the United States, with estimates of 4 million HCV-infected individuals in the United States and 170 million worldwide¹. Most (70–80%) HCV infections persist and about 30% of individuals with persistent infection develop chronic liver disease, including cirrhosis and hepatocellular carcinoma². Epidemiological, viral and host factors have been associated with the differences in HCV clearance or persistence, and studies have demonstrated that a strong host immune response against HCV favours viral clearance^{3,4}. Thus, variation in genes involved in the immune response may contribute to the ability to clear the virus. In a recent genome-wide association study, a single nucleotide polymorphism (rs12979860) 3 kilobases upstream of the *IL28B* gene, which encodes the type III interferon IFN- λ 3, was shown to associate strongly with more than a twofold difference in response to HCV drug treatment⁵. To determine the potential effect of rs12979860 variation on outcome to HCV infection in a natural history setting, we genotyped this variant in HCV cohorts comprised of individuals who spontaneously cleared the virus ($n = 388$) or had persistent infection ($n = 620$). We show that the C/C genotype strongly enhances resolution of HCV infection among individuals of both European and African ancestry. To our

knowledge, this is the strongest and most significant genetic effect associated with natural clearance of HCV, and these results implicate a primary role for *IL28B* in resolution of HCV infection.

Approximately 30% of individuals spontaneously clear acute HCV infection. Host genetic variation is assumed to explain the heterogeneity in HCV clearance across individuals because such differences occur even after exposure to the same HCV inoculum and because there are ethnic differences in clearance frequency^{6,7}. Variation in genes involved in the immune response has already been linked to outcome of acute HCV infection^{8,9}, presumably owing to alteration in the strength and quality of the immune response. However, most variability in spontaneous HCV clearance remains unexplained.

A recent genome-wide association study of >1,600 individuals chronically infected with hepatitis C participating in a clinical treatment trial with pegylated interferon (IFN)- α and ribavirin identified a single nucleotide polymorphism (SNP) on chromosome 19q13, rs12979860, that was strongly associated with sustained virological response (SVR)^{5,10}. This SNP maps 3 kilobases (kb) upstream of the *IL28B* gene, which encodes the type III interferon IFN- λ 3. The C/C genotype was associated with a 2.5 or greater rate (depending on ethnicity) of SVR compared with the T/T genotype, and the C allele was over-represented in a random multi-ethnic population as

Table 1 | Characteristics of study subjects

Characteristic	Clearance ($n = 388$)	Persistence ($n = 620$)
Mean age (years)*	33.9	32.0
Male (%)	78.6 (305)	80.2 (497)
European ancestry (%)	67.3 (261)	61.5 (381)
African ancestry (%)	25 (97)	31.1 (193)
Other (%)	7.7 (30)	7.4 (46)
HBsAg status (% positive)†	9.7 (30)	3.0 (18)
HIV status (% positive)	19.3 (75)	24.4 (151)
rs12979860 allele frequency (%)		
C	European ancestry = 80.3; African ancestry = 56.2	European ancestry = 66.7; African ancestry = 37
T	European ancestry = 19.7; African ancestry = 43.8	European ancestry = 33.3; African ancestry = 63

A total of 68.7% of study subjects was derived from cohorts that were matched on HIV status, gender and ethnicity. Numbers (n) are given in parentheses.

*There was one individual in the clearance group for which age was not available.

†Information for HBsAg (hepatitis B surface antigen) status was unavailable for 80 individuals in the clearance group and 26 in the persistence group.

¹Johns Hopkins University, Division of Infectious Diseases, Baltimore, Maryland 21205, USA. ²Cancer and Inflammation Program, Laboratory of Experimental Immunology, SAIC Frederick, Inc., NCI-Frederick, Frederick, Maryland 21702, USA, and Ragon Institute of Massachusetts General Hospital, Massachusetts Institute of Technology and Harvard University, Boston, Massachusetts 02114, USA. ³Institute for Genome Sciences & Policy, Center for Human Genome Variation, Duke University, Durham, North Carolina 27708, USA. ⁴Department of Genetics, Yale University School of Medicine, 333 Cedar Street, New Haven, Connecticut 06520, USA. ⁵Division of Medicine, Imperial College, London W2 1NY, UK. ⁶Department of Medicine, University of Cambridge, Cambridge CB2 2QQ, UK. ⁷Infections & Immunoepidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, Maryland 20852, USA. ⁸Department of Epidemiology, Johns Hopkins University Bloomberg School of Public Health, Baltimore, Maryland 21205, USA. ⁹Rho, Inc., Chapel Hill, North Carolina 27517, USA. ¹⁰University of Colorado Health Sciences Center, Division of Gastroenterology and Hepatology, Aurora, Colorado 80045, USA. ¹¹Blood Systems Research Institute, San Francisco, California 94118, USA. ¹²Duke Clinical Research Institute and Division of Gastroenterology, School of Medicine, Duke University, Durham, North Carolina 27705, USA. ¹³Ragon Institute of Massachusetts General Hospital, Massachusetts Institute of Technology and Harvard University, Boston, Massachusetts 02114, USA.

*These authors contributed equally to this work.

Table 2 | Effect of *IL28B* rs12979860 genotype on clearance of HCV

Genotype	Frequency of clearance (%)	Frequency of persistence (%)	Comparison	OR (95% CI)	P-value
All subjects*					
T/T	23.4 (37)	76.6 (121)	C/C versus T/T	0.29 (0.18–0.47)	4×10^{-7}
C/T	29.5 (124)	70.5 (297)	C/C versus C/T	0.35 (0.25–0.48)	4×10^{-11}
C/T+T/T	28 (161)	72 (418)	C/C versus C/T+T/T	0.33 (0.25–0.45)	3×10^{-13}
C/C	53 (227)	47 (202)	–	–	–
Subjects of European ancestry†					
T/T	31.4 (16)	68.6 (35)	C/C versus T/T	0.50 (0.25–0.98)	0.04
C/T	27.8 (71)	72.2 (184)	C/C versus C/T	0.36 (0.24–0.52)	1×10^{-7}
C/T+T/T	28.4 (87)	71.6 (219)	C/C versus C/T+T/T	0.38 (0.26–0.54)	1×10^{-7}
C/C	51.8 (174)	48.2 (162)	–	–	–
Subjects of African ancestry†					
T/T	20.8 (20)	79.2 (76)	C/C versus T/T	0.21 (0.10–0.44)	3×10^{-5}
C/T	33 (45)	67 (91)	C/C versus C/T	0.40 (0.21–0.75)	0.005
C/T+T/T	28 (65)	72 (167)	C/C versus C/T+T/T	0.32 (0.17–0.57)	1×10^{-4}
C/C	55.2 (32)	44.8 (26)	–	–	–

OR, odds ratio; CI, confidence interval. Numbers (*n*) are indicated in parentheses for columns two and three.

*OR and *P*-values for all subjects were adjusted by cohort and ethnicity.

†OR and *P*-values for subjects of European and African ancestry were adjusted by cohort.

compared with the chronically infected study cohort, raising the possibility that the C allele may favour spontaneous clearance of HCV.

To address directly the role of the rs12979860 SNP in HCV clearance, we genotyped 1,008 individuals from 6 independent HCV cohorts composed of individuals who cleared virus (*n* = 388) and individuals with persistent infection (*n* = 620). Genotypes were in Hardy–Weinberg equilibrium in both individuals of African and European ancestry (*P* = 0.47 and 0.77, respectively). The frequency of the C allele was significantly greater among individuals of European ancestry than those of African ancestry in both the clearance (*P* = 3×10^{-10}) and persistence groups (*P* = 1×10^{-21}) (Table 1). In both ethnic groups, however, there were significant differences in allele frequencies (C versus T) between the clearance and persistence groups, where the C allele showed greater frequencies in the clearance group than in the persistence group (80.3% versus 66.7% respectively in individuals of European ancestry, *P* = 7×10^{-8} ; 56.2% versus 37% respectively in individuals of African ancestry, *P* = 1×10^{-5}).

More striking differences were observed in an analysis of genotype frequencies where patients with the C/C genotype were three times more likely to clear HCV relative to patients with the C/T and T/T genotypes combined (odds ratio (OR) = 0.33, *P* < 10^{-12} for combined ethnic groups; Table 2 and Fig. 1). Stratification of this analysis by ethnicity indicated that the strength of the protective C/C effect was similar in individuals of African and European ancestry (OR = 0.32, *P* = 1×10^{-4} and OR = 0.38, *P* = 1×10^{-7} , respectively). However, a comparison of the C/C to the T/T group alone suggested stronger protection conferred by C/C in individuals of African ancestry (OR = 0.21, *P* = 3×10^{-5}) relative to that in individuals of European ancestry (OR = 0.50, *P* = 0.04), although our power to detect a true difference is limited owing to small sample sizes in some groups.

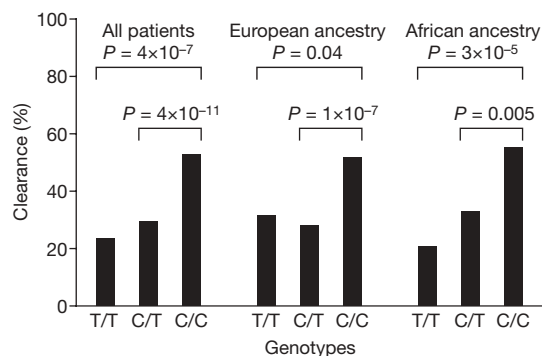


Figure 1 | Percentage of HCV clearance by rs12979860 genotype. Data are shown for all patients, as well as individuals of European ancestry and African ancestry separately.

Table 3 | rs12979860 C allele frequency in worldwide populations

No.	Population	Region	<i>n</i>	C allele frequency (%)
1	Biaka Pygmies	Africa	66	23.5
2	Mbuti Pygmies	Africa*	39	23.1
3	Chagga	Africa*	44	37.5
4	Ethiopian Jews	Africa*	21	54.8
5	Masai	Africa*	20	40.0
6	Sandawe	Africa*	37	44.6
7	Zaramo	Africa	39	37.2
8	Hausa	Africa*	38	31.6
9	Ibo	Africa*	47	38.3
10	Yoruba	Africa*	77	31.2
11	Danish	Europe*	51	76.5
12	Finns	Europe*	33	65.2
13	Hungarians	Europe*	142	65.1
14	Irish	Europe*	113	73.9
15	Russians, Vologda	Europe*	48	61.4
16	Russians	Europe*	32	64.1
17	Adygei	Europe*	53	52.8
18	Chuvash	Europe*	40	73.7
19	Khanty	Europe	49	85.7
20	Komi	Europe*	47	70.2
21	Roman Jews	Europe*	27	79.6
22	Sardinians	Europe	34	52.9
23	European-American	Europe	92	67.4
24	Druze	Southwest Asia	96	77.6
25	Kuwaitis	Southwest Asia	16	75.0
26	Yemenite Jews	Southwest Asia	41	69.5
27	Indians	South Asia*	29	65.5
28	Kachari	South Asia*	17	94.1
29	Thoti	South Asia*	14	89.3
30	Cambodians	Southeast Asia*	24	97.9
31	Laotians	Southeast Asia*	118	93.6
32	Chinese, Taiwan	East Asia*	47	93.6
33	Chinese, San Francisco	East Asia*	59	97.5
34	Hakka	East Asia*	40	95.0
35	Ami	East Asia	40	98.8
36	Atayal	East Asia	40	100.0
37	Japanese	East Asia*	50	91.0
38	Koreans	East Asia*	54	93.5
39	Yakut	East Asia*	50	90.0
40	Micronesians	Oceania*	36	98.6
41	Nasioi	Oceania*	23	100.0
42	Samoans	Oceania*	8	100.0
43	Papua New Guineans	Oceania	22	70.4
44	Pima, Mexico	North America	99	55.5
45	Mayans	North America	52	37.5
46	Muscogees	North America	10	65.0
47	Ticuna	South America	62	20.2
48	Karitiana	South America	54	82.4
49	Surui	South America	47	77.7
50	Guibiba speakers	South America	12	62.5
51	Quechua	South America	22	63.6

*Samples used in *F_{ST}* estimation.

Overall, the protective effect of C seems to be primarily recessive, as no significant difference was observed between the C/T and T/T genotypes in individuals of African ancestry, European ancestry, or combined ethnic groups for clearance of HCV (data not shown), and C/C was consistently protective relative to C/T and/or T/T (Table 2 and Fig. 1). These results mirror the protective effect of the C/C genotype on SVR after HCV treatment observed previously⁵, where the protection conferred by the C allele also seemed to be recessive in both their Caucasian and African-American patients.

Some individuals used in this study of HCV were co-infected with hepatitis B virus (HBV) and/or human immunodeficiency virus (HIV). To eliminate the possibility that co-infection with these viruses might confound the effect of rs12979860 on HCV outcome, analyses were performed using a multivariate model that included hepatitis B surface antigen status as a co-variate or stratifying by HIV status. Neither of these two chronic viral infections altered the effect of this locus on outcome of an acute HCV infection (Supplementary Tables 1 and 2). We also tested whether there were any differences in the effect of the protective rs12979860 C allele as a function of the route of HCV acquisition (plasma products versus injection drug use), but found no significant differences between the two groups (data not shown). Finally, adjusting by other host genetic factors that associate with clearance of HCV did not alter the protection conferred by the C/C genotype (Supplementary Information I).

Patients with lower baseline HCV viral load respond more favourably to interferon- α treatment^{11,12}. However, little is known regarding the impact of viral load during acute infection on spontaneous HCV clearance because very few HCV-infected individuals are identified and studied at this early phase¹³. We reasoned that the mechanism of protection of the C/C genotype might also extend to greater control of viral load in the chronic phase, but there was no correlation between

rs12979860 genotype and viral load (Supplementary Information II and Supplementary Fig. 1). However, differences in viral load assays used in the various cohorts may mask a small correlation, even after conversion to international units.

The frequency of HCV clearance varies markedly across ethnic groups⁹, and differences in allele frequencies for the rs12979860 SNP were observed in the present study (Table 1) and in another study⁵. Indeed, the observation that the C allele is less frequent among individuals of African descent relative to those of European descent might explain, in part, the observed discrepancy in the frequency of viral clearance in these two ethnic groups, where clearance occurs in 36.4% of HCV infections in individuals of non-African ancestry, but only in 9.3% of infections in individuals of African ancestry⁷. To gain a greater insight into the geographic frequency distribution of the protective C allele, we genotyped 2,371 individuals from 51 populations worldwide (Table 3 and Fig. 2a).

The global pattern of allele frequencies shows a striking pattern in which the allele leading to greater natural HCV clearance is nearly fixed throughout east Asia, has an intermediate frequency in Europe, and is the minor allele in Africa (Fig. 2a). A comparison of the rs12979860 allele frequency diversity across 32 world populations (as measured by F_{ST}) with that for 1,062 SNPs typed in these same samples shows that the rs12979860 polymorphism has a greater differential frequency ($F_{ST} = 0.23$) than most of the other polymorphisms (mean $F_{ST} = 0.12$, standard deviation = 0.1), falling within the upper 12.5 percentile of the distribution of F_{ST} values (Fig. 2b). Notably, the high frequencies of the C allele found in north and eastern Asian populations are not reflected in correspondingly high frequencies in their American relatives. Thus, if this locus has been under selection pressure, changes in the selective force that may be dependent on geographical location probably occurred after the

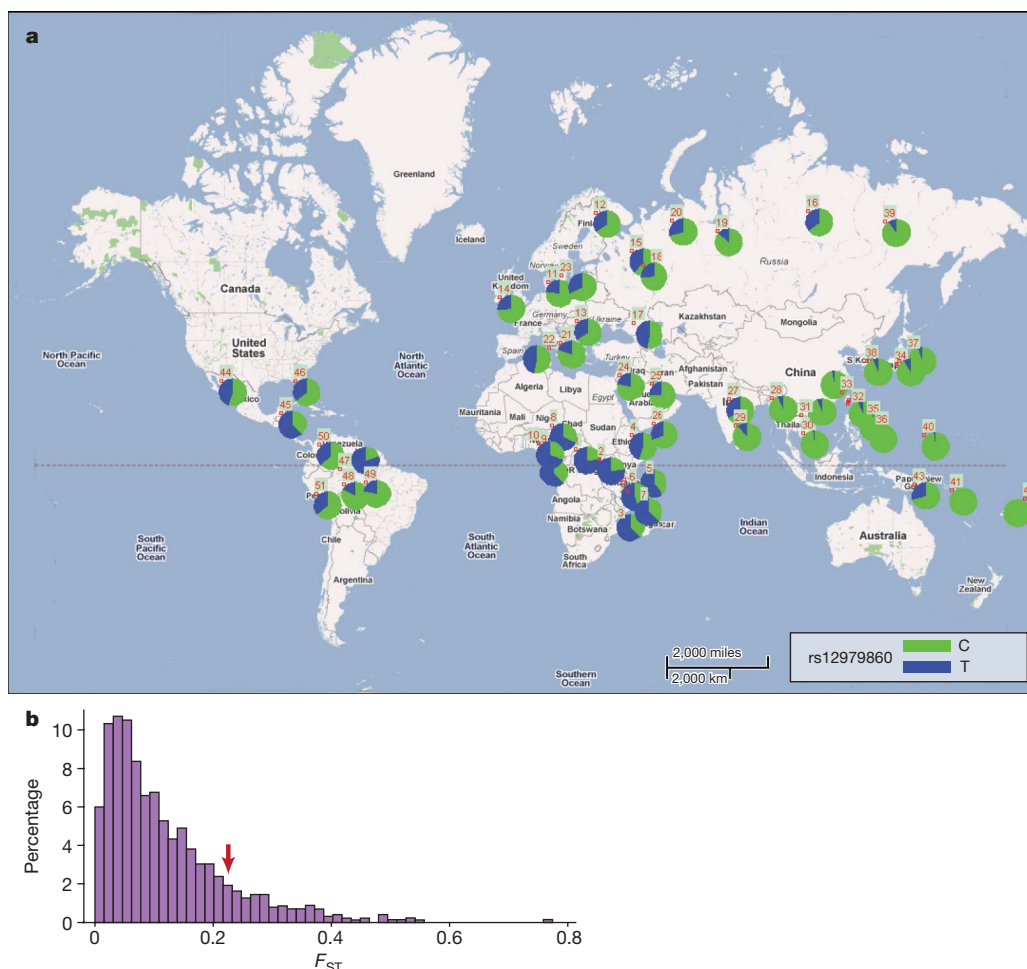


Figure 2 | Sampling locations, allele frequencies and degree of regional differentiation of the rs12979860 C allele. **a**, The numbers identifying populations are given in Table 3. The pie charts show the frequency of the C (green) and T (blue) alleles in each population sampled. **b**, Frequency distribution of F_{ST} values for 1,062 SNPs from 32 of the samples grouped into 6 regions (Africa, Europe, south Asia, southeast Asia, east Asia, Oceania). The red arrow indicates the position of the estimated F_{ST} for rs12979860.

colonization of the New World. That a common variant has such a strong impact on hepatitis C may indicate that it has actually been under selection, adding to an emerging interpretation of genome-wide association studies whereby common variants rarely have large effects unless they were selected to do so¹⁴.

The rs12979860 SNP is only 3 kb upstream of the *IL28B* gene, which encodes the type III interferon IFN- λ 3, and this SNP is in strong linkage disequilibrium ($r^2 > 0.85$) with a non-synonymous coding variant in the *IL28B* gene (213A>G, K70R; rs8103142)⁵. Thus, it is possible that this 213A>G change alters the function of IFN- λ 3 and explains the genetic data described herein, but functional data will be essential to define the precise biological mechanism. Type III interferons include three members—IFN- λ 1, IFN- λ 2 and IFN- λ 3—and the genes encoding these molecules are clustered on human chromosome 19q13 (refs 15, 16). They are structurally related to the IL-10 superfamily of cytokines, but share functional characteristics with the type I interferons (IFN- α and IFN- β) in that they are induced by viral infections, signal through the JAK–STAT pathway, and exhibit antiviral activity *in vitro*^{15,16}. IFN- λ 1 has been shown to exhibit dose- and time-dependent HCV inhibition, induce increases in levels of interferon-stimulated genes, and enhance the antiviral efficacy of IFN- α ¹⁷. It is possible that IFN- λ 3 works through a similar mechanism. *In vitro*, IFN- λ 3 is at least as potent as IFN- λ 1 in terms of protecting HepG2 cells from lysis after infection with encephalomyocarditis virus¹⁸. Severe side effects in HCV treatment have been observed with IFN- α therapy¹⁹, whereas type III IFN (IFN- λ) treatment may exhibit less ‘interferon-like’ adverse effects because receptors for the three family members are only expressed on a limited number of cell types²⁰. Whether IFN- λ may serve as an alternative treatment modality for HCV infection is under investigation.

We have shown that the rs12979860 polymorphism upstream of *IL28B* which was previously associated with HCV treatment response also has a marked impact on natural clearance of HCV and may have been under selection in human history. It is now a priority to determine the mechanisms through which *IL28B* promotes viral defence and the full range of viruses affected by these mechanisms.

METHODS SUMMARY

Subjects in this study were participants in one of six studies: (1) AIDS Link to Intravenous experience (ALIVE²¹; $n = 281$); (2) Multicentre Hemophilia Cohort Study (MHCS²²; $n = 305$); (3) Hemophilia Growth and Development Study (HGDS²³; $n = 106$); (4) Correlates of Resolved Versus Low Level Viremic Hepatitis C Infection in Blood Donors study (REVELL; $n = 85$); (5) an HCV clinic cohort in Portland, Oregon, USA ($n = 51$); and (6) a cohort of injection drug users from the UK ($n = 180$) (see Methods for details). Fifty-one worldwide populations ($n = 2,371$) from the ALlele FREquency Database (ALFRED)²⁴ were also genotyped in this study. Details of sampling and ethnographic information for these populations can be found at <http://alfred.med.yale.edu/>. All populations were in Hardy–Weinberg equilibrium with the exception of one, the Finnish sample ($n = 33$, $P = 0.05$). Genotyping was performed using the ABI TaqMan allelic discrimination kit and the ABI7900HT Sequence Detection System (Applied Biosystems). SAS 9.1 (SAS Institute) was used for statistical analyses.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 16 August; accepted 28 August 2009.

Published online 16 September; corrected 8 October 2009 (see full-text HTML version for details).

- Centers for Disease Control and Prevention. Recommendations for prevention and control of hepatitis C virus (HCV) infection and HCV-related chronic disease. *MMWR Recomm. Rep.* 47, 1–39 (1998).
- Seeff, L. B. Natural history of chronic hepatitis C. *Hepatology* 36 (suppl. 1), S35–S46 (2002).
- Cooper, S. et al. Analysis of a successful immune response against hepatitis C virus. *Immunity* 10, 439–449 (1999).

- Rehermann, B. & Nascimbeni, M. Immunology of hepatitis B virus and hepatitis C virus infection. *Nature Rev. Immunol.* 5, 215–229 (2005).
- Ge, D. et al. Genetic variation in *IL28B* predicts hepatitis C treatment-induced viral clearance. *Nature* advance online publication doi:10.1038/nature08309 (16 August 2009).
- Kenny-Walsh, E. Clinical outcomes after hepatitis C infection from contaminated anti-D immune globulin. Irish Hepatology Research Group. *N. Engl. J. Med.* 340, 1228–1233 (1999).
- Thomas, D. L. et al. The natural history of hepatitis C virus infection: host, viral, and environmental factors. *J. Am. Med. Assoc.* 284, 450–456 (2000).
- Thio, C. L. Host genetic factors and antiviral immune responses to hepatitis C virus. *Clin. Liver Dis.* 12, 713–726 (2008).
- Thio, C. L., Thomas, D. L. & Carrington, M. Chronic viral hepatitis and the human genome. *Hepatology* 31, 819–827 (2000).
- McHutchison, J. G. et al. Peginterferon alfa-2b or alfa-2a with ribavirin for treatment of hepatitis C infection. *N. Engl. J. Med.* 361, 580–593 (2009).
- Lindsay, K. L. et al. A randomized, double-blind trial comparing pegylated interferon alfa-2b to interferon alfa-2b as initial treatment for chronic hepatitis C. *Hepatology* 34, 395–403 (2001).
- Zeuzem, S. et al. Peginterferon alfa-2a in patients with chronic hepatitis C. *N. Engl. J. Med.* 343, 1666–1672 (2000).
- Villano, S. A., Vlahov, D., Nelson, K. E., Cohn, S. & Thomas, D. L. Persistence of viremia and the importance of long-term follow-up after acute hepatitis C infection. *Hepatology* 29, 908–914 (1999).
- Goldstein, D. B. Common genetic variation and human traits. *N. Engl. J. Med.* 360, 1696–1698 (2009).
- Kotenko, S. V. et al. IFN- λ s mediate antiviral protection through a distinct class II cytokine receptor complex. *Nature Immunol.* 4, 69–77 (2003).
- Sheppard, P. et al. IL-28, IL-29 and their class II cytokine receptor IL-28R. *Nature Immunol.* 4, 63–68 (2003).
- Marcello, T. et al. Interferons α and λ inhibit hepatitis C virus replication with distinct signal transduction and gene regulation kinetics. *Gastroenterology* 131, 1887–1898 (2006).
- Dellgren, C., Gad, H. H., Hamming, O. J., Melchjorsen, J. & Hartmann, R. Human interferon- λ 3 is a potent member of the type III interferon family. *Genes Immun.* 10, 125–131 (2009).
- Lauer, G. M. & Walker, B. D. Hepatitis C virus infection. *N. Engl. J. Med.* 345, 41–52 (2001).
- Donnelly, R. P., Sheikh, F., Kotenko, S. V. & Dickensheets, H. The expanded family of class II cytokines that share the IL-10 receptor-2 (IL-10R2) chain. *J. Leukoc. Biol.* 76, 314–321 (2004).
- Vlahov, D. et al. The ALIVE study, a longitudinal study of HIV-1 infection in intravenous drug users: description of methods and characteristics of participants. *NIDA Res. Monogr.* 109, 75–100 (1991).
- Goedert, J. J. et al. A prospective study of human immunodeficiency virus type 1 infection and the development of AIDS in subjects with hemophilia. *N. Engl. J. Med.* 321, 1141–1148 (1989).
- Hilgartner, M. W. et al. Hemophilia growth and development study. Design, methods, and entry data. *Am. J. Pediatr. Hematol. Oncol.* 15, 208–218 (1993).
- Osier, M. V. et al. ALFRED: An allele frequency database for anthropology. *Am. J. Phys. Anthropol.* 119, 77–83 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This project has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health, under Contract number HHSN261200800001E. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government. This Research was supported in part by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research. This research was supported by NIH grants R01DA013324 (D.L.T.), R01DA004334 (G.D.K.), R01HL076902 (M.P.B.), R01DK60590 (H.R.R.) and R01HD41224 (S.M.D.). S.I.K. is a Wellcome Trust Senior Clinical Fellow.

Author Contributions D.L.T., C.L.T., J.G.M., D.B.G. and M.C. designed the study. M.C. directed the study. M.P.M. performed the genotyping. M.P.M. and M.C. wrote the manuscript, with major edits by D.L.T., C.L.T. and D.B.G. The analyses were performed by Y.Q., D.G. and C.O. The samples used in the study were provided by C.L.T., J.K., K.K., S.I.K., G.A., J.J.G., G.D.K., S.M.D., H.R.R., L.H.T., M.P.B. and D.L.T. All authors contributed to preparing the final manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.C. (carrington@mail.nih.gov).

METHODS

Study subjects. The AIDS Link to Intravenous experience (ALIVE) is an ongoing study of injection drug users²¹. The Multicentre Hemophilia Cohort Study (MHCS) is a prospectively followed cohort of patients with coagulation disorders from 16 treatment centres²². The Hemophilia Growth and Development Study (HGDS) is a continuing study of children and adolescents with hemophilia²³. The REVELL study draws both resolved and chronic HCV infections from a large blood bank network consisting of 17 blood centres in the western and southern United States. Individuals with viral clearance in the ALIVE, MHCS and HGDS cohorts were matched to two individuals with viral persistence within the same cohort based on HIV status, gender and ethnicity (African-American, European-American, other). All of the individuals in the HCV clinic cohort cleared the virus, so they were not matched. Participants in the REVELL cohort were all HIV negative and HBV negative. Participants in the UK cohort were individuals referred to hepatology clinics. They were all Caucasian, HBV negative and HIV negative. There was no significant difference in the results when matched and unmatched cohorts were analysed separately (Supplementary Table 3).

HCV testing. HCV infection was established by a second- or third-generation enzyme immunoassay (EIA) (Ortho Diagnostics Systems). Individuals with subsequent negative HCV RNA tests were confirmed anti-HCV positive by a third-generation EIA test or a recombinant immunoblot assay (RIBA) that was separated from the first by a minimum of 6 months. HCV RNA was assessed by a branched DNA (bDNA) assay (Quantiplex HCV RNA 2.0 assay; Chiron Corporation), a qualitative HCV COBAS AMPLICOR system (COBAS AMPLICOR HCV; Roche Diagnostics), or by transcription-mediated amplification (TMA) (Novartis and Gen Probe Inc.). Those subjects with a sample below the limit of detection by the bDNA assay (potential subjects with HCV recovery) had a repeat sample separated by 6 months from the first one tested with the qualitative COBAS. HCV infection

in blood donors (REVELL study) was established by third-generation EIA and confirmed using RIBA. HCV RNA status was established using nucleic acid amplification testing (NAT) of minipools representing 16 donation samples (Procleix HIV-1/HCV Assay, Gen-Probe, Novartis). A reactive minipool result triggered NAT of the individual donations comprising the pool in order to identify the NAT-reactive donation. Residual volume after NAT screening from all antibody positive/RNA negative donors was retested by duplicate undiluted HCV RNA testing using TMA. Individuals with HCV clearance had undetectable HCV RNA in serum or plasma at two time points separated by a minimum of 6 months. Persistently infected individuals had detectable HCV RNA in serum or plasma at two time points separated by a minimum of 6 months.

Statistical analysis. SAS 9.1 (SAS Institute) PROC FREQ was used to compute frequencies and Fisher's exact test *P*-values on categorical variables. PROC LOGISTIC was used to obtain odds ratios and 95% confidence intervals. PROC MEANS was used to calculate mean age. Analyses were performed with all ethnic groups combined and individuals of European and African ancestry separately. All analyses were adjusted for study groups and for ethnicity in three categories (European ancestry, African ancestry and others) when all ethnic groups were combined. Statistical significance refers to two-sided *P*-values of <0.05.

F_{ST} analysis. Population samples were checked in ALFRED to identify all SNPs for which individuals had been typed. For 32 samples an additional 1,062 SNPs were available. Data were pooled by regional affiliation (see Table 3) into six groupings allowing improved frequency estimation: Africa, Europe, south Asia, southeast Asia, east Asia and Oceania. The extent of regional differentiation, F_{ST} , was determined for each of the individual SNPs and its distribution plotted. The F_{ST} of similarly pooled rs12979860 allele frequencies was compared to that found at other loci.

LETTERS

A genome-wide linkage and association scan reveals novel loci for autism

Lauren A. Weiss^{1,2,*†}, Dan E. Arking^{3,*} & The Gene Discovery Project of Johns Hopkins & the Autism Consortium[‡]

Although autism is a highly heritable neurodevelopmental disorder, attempts to identify specific susceptibility genes have thus far met with limited success¹. Genome-wide association studies using half a million or more markers, particularly those with very large sample sizes achieved through meta-analysis, have shown great success in mapping genes for other complex genetic traits. Consequently, we initiated a linkage and association mapping study using half a million genome-wide single nucleotide polymorphisms (SNPs) in a common set of 1,031 multiplex autism families (1,553 affected offspring). We identified regions of suggestive and significant linkage on chromosomes 6q27 and 20p13, respectively. Initial analysis did not yield genome-wide significant associations; however, genotyping of top hits in additional families revealed an SNP on chromosome 5p15 (between *SEMA5A* and *TAS2R1*) that was significantly associated with autism ($P = 2 \times 10^{-7}$). We also demonstrated that expression of *SEMA5A* is reduced in brains from autistic patients, further implicating *SEMA5A* as an autism susceptibility gene. The linkage regions reported here provide targets for rare variation screening whereas the discovery of a single novel association demonstrates the action of common variants.

For a high-resolution genetic study of autism, we selected families with multiple affected individuals (multiplex) from the widely studied Autism Genetic Resource Exchange (AGRE) and US National Institute for Mental Health (NIMH) repositories (Supplementary Methods and Supplementary Table 1). Although the phenotypic heterogeneity in autism spectrum disorders (ASDs) is extensive, in our primary screen we selected families in which at least one proband met Autism Diagnostic Interview-Revised (ADI-R) criteria for diagnosis of autism and included additional siblings in the same nuclear family affected with any autism spectrum disorder. We previously reported an early copy number analysis that revealed a significant role for microdeletion and duplication of 16p11.2 in ASD causation²; here, we present extensive genome-wide linkage and association analyses performed with this high density of SNPs and identify independent and novel genome-wide significant results by both linkage and association analyses.

We combined families and samples from two sources for the primary genetic association screen. The AGRE sample included nearly 3,000 individuals from over 780 multiplex autism families in the AGRE collection³ genotyped at the Broad Institute on the Affymetrix 5.0 platform, which includes over 500,000 SNPs. The NIMH sample included a total of 1,233 individuals from 341 multiplex nuclear families (258 of which were independent of the AGRE sample) genotyped at the Johns Hopkins Center for Complex Disease Genomics on Affymetrix 5.0 and 500K platforms, including the same SNP markers as were genotyped in the AGRE sample.

Before merging, we carefully filtered each data set separately to ensure the highest possible genotype quality for analysis, because technical genotyping artefacts can create false positive findings. We therefore examined the distribution of χ^2 values for the highest quality data, and used a series of quality control (QC) filters designed to identify a robust set of SNPs, including data completeness for each SNP, Mendelian errors per SNP and per family, and a careful evaluation of inflation of association statistics as a function of allele frequency and missing data (see Methods). As 324 individuals were genotyped at both centres, we performed a concordance check to validate our approach. After excluding one sample mix-up, we obtained an overall genotype concordance between the two centres of 99.7% for samples typed on 500K at Johns Hopkins University and 5.0 at the Broad Institute and 99.9% for samples run on 5.0 arrays at both sites. The combined data set, consisting of 1,031 nuclear families (856 with two parents) and a total of 1,553 affected offspring, was used for genetic analyses (Supplementary Table 1). These data were publicly released in October 2007 and are directly available from AGRE and NIMH.

For linkage analyses, the common AGRE/NIMH data set was further merged with Illumina 550K genotype data generated at the Children's Hospital of Philadelphia (CHOP) and available from AGRE, adding ~300 nuclear families (1,499 samples). We used the extensive overlap of samples between the AGRE/NIMH and the CHOP data sets (2,282 samples) to select an extremely high quality set of SNPs for linkage analysis. Specifically, we only included SNPs genotyped in both data sets with >99.5% concordance and ≤ 1 Mendelian error.

Linkage analysis involving high densities of markers, where clusters of markers are in linkage disequilibrium (LD), can falsely inflate the evidence for genetic sharing among siblings when neither parent is genotyped⁴. To alleviate these concerns, we analysed a pruned set of 16,311 highly polymorphic, high-quality autosomal SNPs which were filtered to remove any instances in which two nearby markers were correlated with $r^2 > 0.1$, providing a marker density of ~0.25 cM (see Methods). In this analysis of 878 families, four genomic regions showed LOD scores in excess of 2.0 and one region, 20p13, exceeded the formal genome-wide significance threshold of 3.6 (ref. 5) (maximum LOD, 3.81; Fig. 1a and Supplementary Table 2). Restricting analysis to only those families with both parents genotyped (784 families) showed that these results are not an artefact of missing parental data (Fig. 1b). We further tested the stability of these results by varying the recombination map and halving the marker density by placing every other marker into two non-overlapping SNP sets (Methods Summary); all analyses showed consistent and strong linkage to the same regions (data not shown).

We used the transmission disequilibrium test (TDT) across all SNPs passing quality control in the complete family data set for association

¹Center for Human Genetic Research, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02114, USA. ²Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. ³Center for Complex Disease Genomics, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, Maryland 21205, USA. [†]Present address: Department of Psychiatry, Institute for Human Genetics, Center for Neurobiology and Psychiatry, UCSF, San Francisco, California 94143, USA.

*These authors contributed equally to this work.

‡Lists of participants and affiliations appear at the end of the paper.

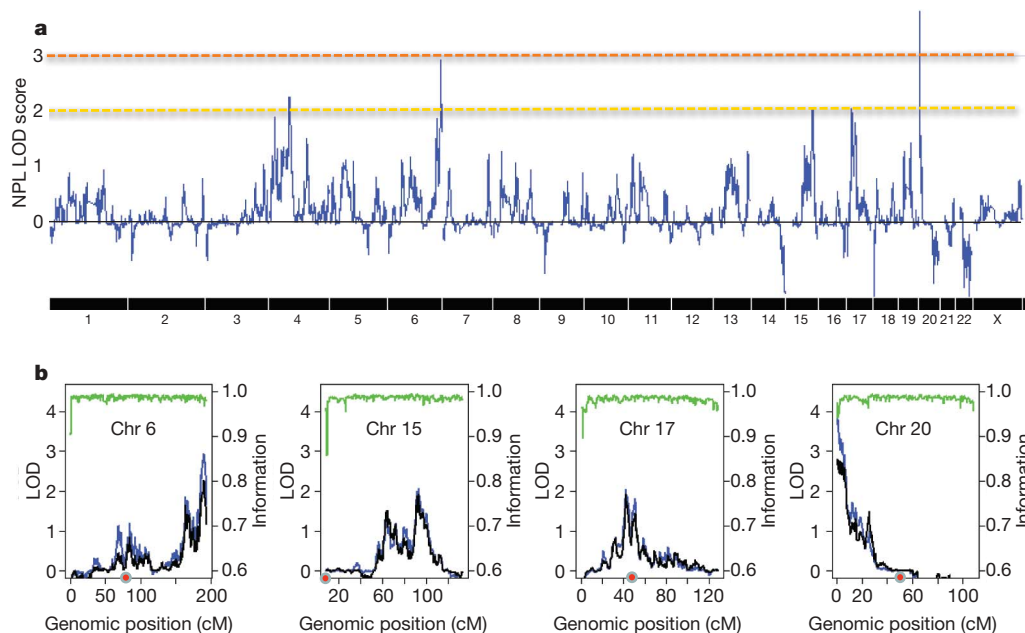


Figure 1 | Genome-wide linkage results. **a**, The genome-wide linkage results are shown, with the orange line indicating non-parametric linkage (NPL) LOD = 3 and the yellow line indicating NPL LOD = 2. **b**, Four chromosomes with LOD > 2. The black and blue lines indicate results from

families with both parents genotyped and all families, respectively. The green line indicates information content (right-hand y axis). The red circle indicates the position of the centromere.

analyses as the TDT is not biased by population stratification. We estimated a threshold for genome-wide significance using both permutation ($P < 2.5 \times 10^{-7}$) and estimating the effective number of tests ($P < 3.4 \times 10^{-7}$), and use the more conservative here (see Methods). No SNP met criteria for genome-wide significance at $P < 2.5 \times 10^{-7}$. However, we observed an excess of independent regions associated at $P < 10^{-5}$ (6 observed versus 1 expected) and $P < 10^{-4}$ (30 observed versus 15 expected) despite the lack of overall statistical inflation ($\lambda = 1.03$, Supplementary Fig. 1), suggesting that common variants in autism exist, but that our initial scan did not have sufficient statistical power to identify them definitively (Table 1 and Supplementary Fig. 2).

For the TDT associations with $P < 10^{-4}$, we additionally used the cases that were excluded from the TDT due to missing parental data. We matched 90 independent and unrelated cases with 1,476 NIMH control samples genotyped on the Affymetrix 500K arrays⁶, and performed case-control association analysis (Supplementary Table 3), combining

these results with the TDT data. Promisingly, we now observed eight SNPs (in seven independent regions) with association at $P < 10^{-5}$ (Table 1). Of note, comparing Caucasian with non-Caucasian samples in the AGRE/NIMH data set, we did not observe significant heterogeneity for top results.

Our strongest associations were at chromosome 4q13 (rs17088254, $P = 8.5 \times 10^{-6}$) between *CENPC1*, a centromere autoantigen, and *EPHA5*, an ephrin receptor potentially involved in neurodevelopment; at 5p15 (rs10513025, $P = 1.7 \times 10^{-6}$) in the EST DB512398, located between *SEMA5A* and *TAS2R1*; at 6p23 (rs7766973, $P = 6.8 \times 10^{-7}$) in *JARID2*, an orthologue of the mouse jumonji gene, encoding a nuclear protein essential for embryogenesis, especially neural tube formation; at 9p24 (rs4742409, $P = 7.9 \times 10^{-6}$) between *PTPRD*, a protein tyrosine phosphatase involved in neurite outgrowth, and *JMJD2C* (also called *KDM4C*), a jumonji-domain containing protein involved in tri-methyl-specific demethylation; at 9q21 (rs952834, $P = 7.8 \times 10^{-6}$) between *ZCCHC6*, a zinc finger and

Table 1 | Top TDT results and replication data

Locus				Scan					Replication				Meta-analysis	
Chromosome	Position	SNP	LD, proxy	T	U	OR	P	P (with case-control)	T	U	OR	P (1-sided)	P (meta)	P (proxy)
4	68019960	rs17088254	–	137	219	0.63	1.4×10^{-5}	8.5×10^{-6}	48	38	1.3	NA	0.011	4.8×10^{-3}
4	68189460	rs2632453	$r^2 = 0.67$, rs17088254	171	245	0.70	2.9×10^{-4}	2.4×10^{-4}	248	234	1.06	NA	0.022	–
5	9676622	rs10513025	–	84	152	0.55	9.6×10^{-6}	1.7×10^{-6}	152	199	0.76	6.1×10^{-3}	2.1×10^{-7}	–
6	15365718	rs13208655	$r^2 = 0.74$, rs7766973	NA	NA	NA	NA	NA	829	831	1.00	0.48	0.48	–
6	15376030	rs7766973	–	631	811	0.78	2.1×10^{-6}	6.8×10^{-7}	139	142	0.98	0.43	2.0×10^{-4}	2.8×10^{-4}
9	7763723	rs4742408	–	591	739	0.80	4.9×10^{-5}	2.7×10^{-4}	241	224	1.08	NA	0.030	–
9	7764180	rs4742409	–	499	645	0.77	1.6×10^{-5}	7.9×10^{-6}	77	87	0.89	0.22	1.6×10^{-4}	3.6×10^{-4}
9	7764774	rs6477233	$r^2 = 0.6$, rs4742409	NA	NA	NA	NA	NA	734	752	0.98	0.32	0.32	–
9	86471331	rs952834	–	656	825	0.80	1.1×10^{-5}	7.8×10^{-6}	173	160	1.08	NA	5.4×10^{-3}	–
10	68842909	rs7923367	–	89	160	0.56	6.8×10^{-6}	3.4×10^{-6}	18	25	0.72	0.14	4.1×10^{-5}	–
11	22775950	rs12293188	–	449	327	1.37	1.2×10^{-5}	1.1×10^{-6}	486	513	0.95	NA	3.0×10^{-3}	–
11	22785182	rs16910190	–	421	308	1.37	2.8×10^{-5}	1.4×10^{-5}	55	67	0.82	NA	0.014	–
11	22785488	rs16910194	–	444	330	1.35	4.2×10^{-5}	3.7×10^{-6}	80	75	1.07	0.34	2.8×10^{-4}	–
11	22791645	rs3763947	–	429	320	1.34	6.8×10^{-5}	3.4×10^{-5}	57	57	1.00	NA	2.4×10^{-3}	–

Top results from the combined TDT and case-control analysis are shown ($P < 10^{-5}$), with replication data, where it exists. For Sequenom genotyping that used a proxy SNP, that SNP and its LD (r^2) with the SNP of interest is shown. Transmitted (T) and untransmitted (U) counts and odds ratios (OR) for the minor allele are shown for each SNP. Replication results are shown for additional autism family data using Affymetrix and Sequenom genotyping technology. The meta-analysis P -value is shown as is the P -value for meta-analysis where proxy SNP data was included. Bold font: $P < 10^{-5}$ TDT/case-control analysis, $P < 0.05$ replication, $P < 2.5 \times 10^{-7}$ meta-analysis. NA, not applicable.

CCHC domain containing protein, and *GAS1*, growth-arrest-specific protein; at 10q21 (rs7923367, $P = 3.4 \times 10^{-6}$) in *CTNNA3*, $\alpha 3$ catenin, which may be involved in the formation of stretch-resistant cell-cell adhesion complexes; and two SNPs on 11p14 (rs12293188, $P = 1.1 \times 10^{-6}$; rs16910194, $P = 3.7 \times 10^{-6}$) in *GAS2*, a caspase-3 substrate that has a role in regulating microfilament and cell shape changes during apoptosis and can modulate cell susceptibility to p53-dependent apoptosis by inhibiting calpain activity (Table 1).

To confirm whether any of these top results might indicate true susceptibility loci, we attempted to replicate these signals, as well as others with $P < 10^{-4}$ in the initial TDT that met stringent genotyping quality criteria (Supplementary Table 3). We used several data sources to replicate the association results. First, we used additional autism family samples (318 trios collected by investigators of the Autism Consortium and in Montreal) with genome-wide Affymetrix 5.0/500K array data also genotyped at the Genetic Analysis Platform of the Broad Institute using the same conditions, QC and analysis pipelines (Methods).

Second, independent Autism Genome Project (AGP) families, along with a set of Finnish families and a set of Iranian trios, were used for replication of our top findings ($n = 1,755$ trios). Two Sequenom replication pools were designed, attempting to include as many of the regions associated at $P < 10^{-4}$ as possible. The full set of SNPs considered and those successfully genotyped are shown in Supplementary Table 3, with linkage disequilibrium (r^2) noted for SNPs selected as proxies for Affymetrix markers. One of the eight SNPs with $P < 10^{-5}$ (rs10513025) that failed in this Sequenom assay was subsequently replaced in a subset of AGP samples with a TaqMan assay. This assay showed 99.89% concordance with Affymetrix genotypes in the overlapping AGRE-NIMH samples (2,797 out of 2,800 concordant genotypes), with manual review of the Affymetrix genotype calls also confirming the marker to be of extremely high quality (Supplementary Fig. 4). In the independent replication effort, only rs10513025 was associated with $P < 0.01$ (Table 1).

Combining the scan and replication data, only rs10513025 met criteria for genome-wide significance defined by LD and permutation analyses ($P < 2.5 \times 10^{-7}$). To increase coverage of this region and fill in missing genotypes and SNPs that failed quality control, we performed imputation analysis. rs10513026 was highly (but not perfectly) correlated to the replicated chromosome 5 SNP (rs10513025) and showed even stronger association than originally observed with rs10513025 (Supplementary Fig. 3). These and several other promising SNPs were directly genotyped in the original scan samples and, in fact, showed higher levels of significance (Table 2). Direct genotyping confirmed that rs10513026 showed stronger association than rs10513025 (P -value 4.5×10^{-6} versus 9.8×10^{-6} in the re-genotyped scan trios), increasing the significance of this observation further. Several other promising results from this analysis were genotyped in a subset of scan samples, and, of note, the top SNP in imputation analysis (rs10874241, imputation $P = 9.8 \times 10^{-7}$, odds ratio (OR) = 0.43) showed consistent results (OR = 0.4, $P = 4 \times 10^{-7}$) when directly genotyped (Supplementary Table 4).

rs10513025 and neighbours are on chromosome 5p15 in a region of LD containing several other ESTs and *TAS2R1*, a bitter taste receptor (Supplementary Fig. 3). The SNPs are ~80 kb upstream of semaphorin 5A (*SEMA5A*), a gene implicated in axonal guidance and known to be downregulated in lymphoblastoid cell lines of autism

Table 2 | Chromosome 5p15 SNPs

SNP	Chromosome	Position	MAF	OR	P	Replication P
rs10513025	5	9676622	0.041	0.5526	9.58×10^{-6}	0.006059
rs10513026	5	9677106	0.040	0.53	4.50×10^{-6}	NA
rs16883317	5	9701592	0.038	0.53	7.20×10^{-5}	NA

Three SNPs in the chromosome 5p15 association locus genotyped by Sequenom iPLEX are shown, with minor allele frequency (MAF), odds ratio (OR) and P -value in the AGRE and NIMH sample, as well as replication data from all available samples for rs10513025 (see Methods). NA, not applicable.

cases versus healthy controls⁷. An independent study at Children's Hospital Boston using whole blood (S.W.K., L.K. and Z.K., manuscript in preparation) confirms this lower expression ($P = 0.0034$) of *SEMA5A* in autism cases versus controls. To evaluate the role of this locus in autism pathogenesis more completely, we evaluated the entirety of 5p15 for copy-number variation. Despite excellent probe coverage throughout the locus, no common or rare copy number variants were detected in the entire AGRE scan in the region of LD surrounding the associated SNPs and the entire *SEMA5A* locus including 250 kb up- and downstream (see Methods).

To test directly *SEMA5A* expression in brains from autistic patients, tissue samples from 20 cases with a primary diagnosis of autism and 10 controls were obtained through the Autism Tissue Program and the Harvard Brain Bank. Samples were dissected from Brodmann area 19 of the occipital lobe cortex, a region demonstrating differences between autism cases and controls in functional imaging studies, and subjected to quantitative PCR⁸. *SEMA5A* expression, determined relative to *MAP2* (neuron specific), was significantly lower in autism brains than controls after adjustment for the age at brain acquisition, post-mortem interval and sex ($P = 0.024$, Fig. 2).

We also analysed our data for association signals at candidate genes or regions with previous evidence of involvement in autism. Although there are few well-replicated associations of biological candidate genes, there are many rare genetic variants, diseases and syndromes associated with autism. Most of these loci have not been systematically assessed to see whether common variation in the gene or region might contribute to autism. We assessed four categories of candidate loci: (1) genes with previous evidence for association with common variation; (2) genes implicated by rare variants leading to autism; (3) genes causing Mendelian diseases associated with autism; and (4) regions where microdeletion or microduplication syndromes are associated with autism. For each gene, we included all SNPs passing basic quality criteria within 2 kb of the transcript.

Overall, there were no compelling results in these sets (all $P > 10^{-4}$), considering the number of SNPs tested, and only two regions met criteria for region-wide (only SNPs in that gene/region considered) or set-wide (for example, all candidate regions in the set of common variant genes considered) significance by permutation testing (Supplementary Table 5). *MECP2* (Rett syndrome) met criteria for region-wide association ($P = 0.0071$, 5 SNPs, Supplementary Table 5). Moreover, the Williams syndrome region was borderline for set-wide significance ($P = 0.051$, Supplementary Table 5). One SNP in particular showed strong association (rs2267831, $P = 0.00012$, OR = 0.56)—as this was a rare SNP with undertransmission of the minor allele, we genotyped a subset of families and observed similar, slightly less significant distortion (OR = 0.61). The SNP is located within *GTF2IRD1*, a transcription factor within the critical region for the Williams syndrome cognitive behavioural profile^{9–11}.

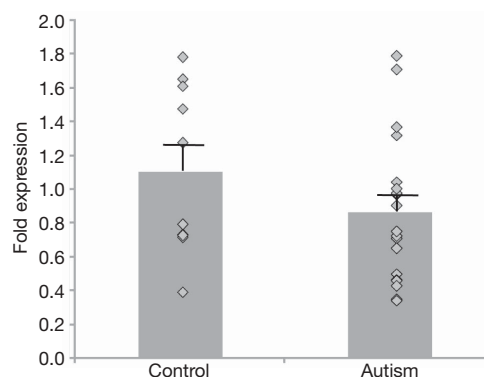


Figure 2 | *SEMA5A* expression in autism brains. *SEMA5A* gene expression is shown relative to *MAP2*. Diamonds indicate individual expression levels for each sample; error bars indicate standard error (s.e.).

There seems to be little overlap between the regions of strongest linkage and association in this study. A more detailed assessment of SNP and haplotype association in the most significant linkage regions did not yield common variation that could explain the evidence for linkage (Supplementary Table 6). This is an expected outcome if linkage signals arise from rare, high penetrance variation (for which the genotyping arrays do not offer an adequate proxy) whereas association is sensitive to common variation with lower penetrance (that cannot be detected by linkage). For example, a 0.3% variant that increases risk by tenfold would readily be picked up by this informative linkage scan, but would very likely not be assessed by the common SNPs on the Affymetrix 5.0 array; by contrast, the modest and protective impact of the 5% variant at the *SEMA5A* rs10513025 creates no detectable excess allele sharing among siblings but is strongly detected by association.

During review of this manuscript, another genome-wide association study (GWAS) was published which identified significant association to SNPs on chromosome 5p14¹². Although there was significant overlap between study samples, each of these scans contained a large set of unique families, so we sought to evaluate independent evidence of the top SNP (rs4307059) reported at 5p14. This SNP happens to be directly genotyped by both Affymetrix and Illumina platforms. We have a sizable number ($n = 796$) of affected subjects with two parents genotyped (and of predominantly similar European background). However, we observed no support for association at this locus (T:U 354:335 in favour of the minor allele, a trend in the opposite direction as reported).

Autism genes have been difficult to identify, despite the high heritability of autism spectrum disorders. Up to 10% of autism cases may be due to rare sequence and gene dosage variants, for example, mutations in *NRXN1*, *NLGN3/NLGN4*, *SHANK3* and copy number variants at 15q11-q13 and 16p11.2. A number of diseases of known aetiology, including Rett syndrome, fragile X syndrome, neurofibromatosis type I, tuberous sclerosis, Potocki-Lupski syndrome, and Smith-Lemli-Opitz syndrome are also associated with autism^{1,13}. However, the remaining 90% of autism spectrum disorders, although highly familial, have unknown genetic aetiology. A genome-wide linkage study using the Affymetrix 10K SNP array to genotype over 1,000 families found no genome-wide significant linkage signals, but documented suggestive linkage at 11p12-p13 and 15q23-q25 and reinforced a modest role for rare copy-number variants¹⁴.

Many complex diseases have recently had great success with GWAS approaches, but most identified modest effects with odds ratios less than 1.3 (<http://www.genome.gov/26525384>). Our association analysis has excellent statistical power ($>80\%$) to find effects of relatively common alleles (0.01–0.25 in frequency) explaining 1% of the variance in autism at the genome-wide significant level. It is near-perfectly powered for alleles of SNPs present on the array (or perfectly proxied) down to 1% at the replication cutoff $P < 10^{-4}$, assuming additive background genetic variance of 0.8 and shared environmental variance of 0.05 with prevalence of 0.006. One of the advantages of a family-based association test is that we avoid false positive results generated by population stratification, and in addition, we have performed careful quality control to reduce the chances of being misled by technical artefacts. However, the SNP coverage of the Affymetrix 5.0 chips is incomplete; in fact, a recent re-sequencing survey suggests that these arrays assay only 57% of variants with minor allele frequency (MAF) $>5\%$ at $r^2 = 0.8$ (ref. 15). We therefore cannot exclude untested variation of large effect in autism. The linkage analysis, assuming a fully informative marker in 800 sibling pairs, should detect sibling allele sharing of at least 55.125%¹⁶.

Our linkage analysis revealed two novel regions of linkage, 6q27 (LOD = 2.94) and 20p13 (LOD = 3.81), with the latter formally exceeding the threshold for genome-wide significance. There is some overlap between the more modest signals (LOD >2 on chromosome 15 and chromosome 17) and previously reported suggestive linkage signals, but little overlap with the most promising regions of common

SNP association. This suggests that the regions of the genome showing linkage may harbour rare variation, potentially with allelic heterogeneity across families, which would require re-sequencing to uncover, as has been demonstrated for the 7q35 region^{17–19}. Interestingly, several of these regions overlap with rare syndromes or genetic events known to be strong risk factors for autism. For example, an autism case with a translocation disrupting 15q25 has been reported, whereas the 17p region overlaps the Smith-Magenis and Potocki-Lupski syndrome region.

The initial TDT analysis of this large multiplex autism data set did not reveal any associations meeting criteria for genome-wide significance, suggesting that there are not many common loci of moderate to large effect size even in a highly heritable disorder like autism. Nevertheless, replication data in our study identified a novel locus with genome-wide significant evidence for association to autism. In addition, several other SNPs in the region show similarly strong association (rs10513026, rs16883317). We ascertained a large replication sample from independent family studies with a replication at $P = 0.0061$ and meta-analysis showed this association ($P = 2.12 \times 10^{-7}$) to meet criteria for genome-wide association in our experiment. This region on chromosome 5 harbours the gene encoding the bitter taste receptor, *TAS2R1*, and several uncharacterized ESTs and is adjacent to *SEMA5A*, a member of the semaphorin axonal guidance protein family, which has shown downregulated expression in transformed B lymphocytes from autism samples⁷. We have further extended this finding by directly demonstrating lowered *SEMA5A* gene expression in autism brain tissue. This is an attractive candidate gene given that its protein is a bi-functional guidance molecule, which is both attractive and inhibitory for developing neurons. Interestingly, the *SEMA5A* receptor is plexin B3, which also signals through the tyrosine kinase MET, a previously reported autism susceptibility gene^{20,21}.

Finally, we investigated whether different classes of genes or regions—loci previously implicated by functional or positional candidate gene association studies, rare variants implicated in autism, Mendelian disorder genes with association to autism, or regions of copy number variation associated with autism—showed association with common alleles included in our marker set. Although there were several nominally significant associations, only the Williams syndrome region (one SNP in *GTF2IRD1*) was borderline statistically significant ($P = 0.051$), after correcting for the microdeletion/duplication syndrome regions tested. In the category of Mendelian disorders associated with autism, *MECP2*, the gene for Rett syndrome, showed region-wide statistical significance. These results raise the possibility that Rett and Williams syndrome genes may contribute more generally to autism spectrum disorders. Although the genes in which common variation has been reported to be associated with autism do not show evidence for association, this cannot be interpreted as failure to replicate previous results in all cases, because much of the variation reported as associated is not captured on the Affymetrix platform (for example, length polymorphisms, microsatellites, untagged SNPs such as the promoter variant at *MET*²¹). Instead, despite a high density of markers, our results suggest that we did not identify additional common variation with evidence for association. Overall however, our results indicate that these postulated candidate regions, mostly based on rare events known to cause autism, are not among the regions with common alleles having the strongest risk effects for autism.

Interestingly, both our linkage and association analyses, from the primary and replication analyses, suggest that low-frequency (<0.05) minor alleles may be common in autism. Intriguingly, the linkage studies reveal low-frequency susceptibility alleles whereas the association analyses have uncovered rare alleles with odds ratios less than 0.6 (the common alleles in the population associated with increased risk for autism). This can occur when the ancestral allele, that was previously neutral or beneficial, now has detrimental effects revealed by an evolutionarily recent environment, or when a pleiotropic function of the allele is selectively advantageous, or when this variation is

hitch-hiking on a shared haplotype with a distinct beneficial allele²². However, it is worth noting that our study design of ascertaining multiplex families is not well powered to identify loci under this genetic model of common major alleles associated with autism susceptibility.

We report genome-wide significant linkage as well as an association of common genetic variation with autism. Our results will require follow-up to identify the functional variation in the linkage and association regions that we report here and to probe the functions of the relatively unstudied transcripts implicated. These results could provide completely novel insight into the biology and pathogenesis of a common neurodevelopmental disorder.

METHODS SUMMARY

Samples and genotyping. Our primary samples are from the AGRE and NIMH Repositories. Replication with Affymetrix technology included NIMH controls, families collected by members of the Autism Consortium, and families ascertained from Montreal. Replication with Sequenom technology included the Autism Genome Project, Finnish, and Iranian subsets of Autism Consortium investigator-collected families. Details of the ascertainment for each sample collection, genotyping and quality control processes can be found in Methods. **Linkage and association analysis.** The linkage analysis was conducted with a pruned autosomal SNP set (see Methods for details of marker selection) and chromosome X set (670 SNPs) using the cluster option in MERLIN/MINX ($r^2 < 0.1$)²³, yielding 16,581 independent markers. We performed confirmatory analysis on non-overlapping data sets by selecting alternative SNPs.

Association analysis was performed in PLINK²⁴. The basic association test was a transmission disequilibrium test (TDT), and the extra cases versus controls analysis was performed by allelic association, after excluding cases that were not well matched to the controls, based on multi-dimensional scaling ($\lambda < 1.1$). Combining the TDT and case-control tests was performed using expected and observed allele counts by the formula $Z_{\text{meta}} = (\sum \text{exp} - \sum \text{obs}) / \sqrt{\sum \text{var}}$. Meta-analysis of AGRE/NIMH and replication data was performed using the statistic $(Z_{\text{AGRE/NIMH}} + Z_{\text{replication}}) / \sqrt{2}$. Gene-set analysis was performed in PLINK using the set-based TDT. Imputation-based association was performed in PLINK with the proxy-tdt command, using the HapMap CEU parent samples as the reference panel and information score > 0.8 . Haplotype analysis in the linkage regions was performed using 5-SNP sliding windows, as implemented in PLINK hap-tdt. See Methods for details of determination of genome-wide significance thresholds.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 22 August 2008; accepted 8 September 2009.

1. Abrahams, B. S. & Geschwind, D. H. Advances in autism genetics: on the threshold of a new neurobiology. *Nature Rev. Genet.* **9**, 341–355 (2008).
2. Weiss, L. A. *et al.* Association between microdeletion and microduplication at 16p11.2 and autism. *N. Engl. J. Med.* **358**, 667–675 (2008).
3. Geschwind, D. H. *et al.* The autism genetic resource exchange: a resource for the study of autism and related neuropsychiatric conditions. *Am. J. Hum. Genet.* **69**, 463–466 (2001).
4. Abecasis, G. R. & Wigginton, J. E. Handling marker-marker linkage disequilibrium: pedigree analysis with clustered markers. *Am. J. Hum. Genet.* **77**, 754–767 (2005).
5. Lander, E. & Kruglyak, L. Genetic dissection of complex traits: Guidelines for interpreting and reporting linkage results. *Nature Genet.* **11**, 241–247 (1995).
6. Sklar, P. *et al.* Whole-genome association study of bipolar disorder. *Mol. Psychiatry* **13**, 558–569 (2008).
7. Melin, M. *et al.* Constitutional downregulation of SEMA5A expression in autism. *Neuropsychobiology* **54**, 64–69 (2006).
8. Gaffrey, M. S. *et al.* Atypical participation of visual cortex during word processing in autism: an fMRI study of semantic decision. *Neuropsychologia* **45**, 1672–1684 (2007).
9. Hirota, H. *et al.* Williams syndrome deficits in visual spatial processing linked to GTF2IRD1 and GTF2I on chromosome 7q11.23. *Genet. Med.* **5**, 311–321 (2003).
10. Edelmann, L. *et al.* An atypical deletion of the Williams-Beuren syndrome interval implicates genes associated with defective visuospatial processing and autism. *J. Med. Genet.* **44**, 136–143 (2007).
11. van Hagen, J. M. *et al.* Contribution of CYLN2 and GTF2IRD1 to neurological and cognitive symptoms in Williams Syndrome. *Neurobiol. Dis.* **26**, 112–124 (2007).
12. Wang, K. *et al.* Common genetic variants on 5p14.1 associate with autism spectrum disorders. *Nature* **459**, 528–533 (2009).
13. Zafeiriou, D. I., Ververi, A. & Vargiami, E. Childhood autism and associated comorbidities. *Brain Dev.* **29**, 257–272 (2007).
14. Szatmari, P. *et al.* Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nature Genet.* **39**, 319–328 (2007).

15. Bhangale, T. R., Rieder, M. J. & Nickerson, D. A. Estimating coverage and power for genetic association studies using near-complete variation data. *Nature Genet.* **40**, 841–843 (2008).
16. Risch, N. J. Searching for genetic determinants in the new millennium. *Nature* **405**, 847–856 (2000).
17. Arking, D. E. *et al.* A common genetic variant in the neurexin superfamily member CNTNAP2 increases familial risk of autism. *Am. J. Hum. Genet.* **82**, 160–164 (2008).
18. Alarcón, M. *et al.* Linkage, association, and gene-expression analyses identify CNTNAP2 as an autism-susceptibility gene. *Am. J. Hum. Genet.* **82**, 150–159 (2008).
19. Bakkaloglu, B. *et al.* Molecular cytogenetic analysis and resequencing of contactin associated protein-like 2 in autism spectrum disorders. *Am. J. Hum. Genet.* **82**, 165–173 (2008).
20. Campbell, D. B. *et al.* Disruption of cerebral cortex MET signaling in autism spectrum disorder. *Ann. Neurol.* **62**, 243–250 (2007).
21. Campbell, D. B. *et al.* A genetic variant that disrupts MET transcription is associated with autism. *Proc. Natl Acad. Sci. USA* **103**, 16834–16839 (2006).
22. Di Rienzo, A. Population genetics models of common diseases. *Curr. Opin. Genet. Dev.* **16**, 630–636 (2006).
23. Abecasis, G. R., Cherny, S. S., Cookson, W. O. & Cardon, L. R. Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. *Nature Genet.* **30**, 97–101 (2002).
24. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank all of the families who have participated in and contributed to the public resources that we have used in these studies. The Broad Institute Center for Genotyping and Analysis is supported by grant U54 RR020278 from the National Center for Research Resources. The Gene Discovery Project of Johns Hopkins was funded by grants from the National Institutes of Mental Health (MH60007, MH081754) and the Simons Foundation. This study was funded in part through a grant from the Autism Consortium of Boston. Support for the Extreme Discordant Sib-Pair (EDSP) family sample was provided by the NLM Family foundation. Support for the Massachusetts General Hospital (MGH)—Finnish collaborative sample was provided by NARSAD. Support for the Homozygosity Mapping Collaborative for Autism (HMCA) came from NIMH (1R01 MH083565), the Nancy Lurie Marks (NLM) Family Foundation and the Simons Foundation. Eric M. Morrow is supported by the NIMH (1K23MH080954). Support for the Iranian family sample was provided by the Special Education Organization of Iran, under the Iranian Ministry of Education. Lauren A. Weiss was supported by a Ruth L. Kirschstein National Research Service Award and is currently supported by the International Mental Health Research Organization. The collection of data and biomaterials that participated in the National Institute of Mental Health (NIMH) Autism Genetics Initiative has been supported by National Institute of Health grants MH52708, MH39437, MH00219 and MH00980; National Health Medical Research Council grant 0034328; and by grants from the Scottish Rite, the Spunk Fund, Inc., the Rebecca and Solomon Baker Fund, the APEX Foundation, the National Alliance for Research in Schizophrenia and Affective Disorders (NARSAD), and the endowment fund of the Nancy Pritzker Laboratory (Stanford); and by gifts from the Autism Society of America, the Janet M. Grace Pervasive Developmental Disorders Fund, and families and friends of individuals with autism. The NIMH collection Principal Investigators and Co-Investigators were: Neil Risch, Richard M. Myers, Donna Spiker, Linda J. Lotspeich, Joachim F. Hallmayer, Helena C. Kraemer, Roland D. Ciaranello, Luigi Luca Cavalli-Sforza (Stanford University, Stanford); William M. McMahon and Brent Petersen (University of Utah, Salt Lake City). The Stanford team is indebted to the parent groups and clinician colleagues who referred families and extends their gratitude to the families with individuals with autism who were partners in this research. The collection data and biomaterials also come from the Autism Genetic Resource Exchange (AGRE) collection. This program has been supported by a National Institute of Health grant MH64547 and the Cure Autism Now Foundation. The AGRE collection Principal Investigator is Daniel H. Geschwind (UCLA). The Co-Principal Investigators include Stanley F. Nelson and Rita M. Cantor (UCLA), Christa Lese Martin (Univ. Chicago), T. Conrad Gilliam (Columbia). Co-Investigators include Maricela Alarcon (UCLA), Kenneth Lange (UCLA), Sarah J. Spence (UCLA), David H. Ledbetter (Emory) and Hank Juo (Columbia). Scientific oversight of the AGRE program is provided by a steering committee (Chair: Daniel H. Geschwind; Members: W. Ted Brown, Maja Bucan, Joseph D. Buxbaum, T. Conrad Gilliam, David Greenberg, David H. Ledbetter, Bruce Miller, Stanley F. Nelson, Jonathan Pevsner, Carol Sprouse, Gerard D. Schellenberg and Rudolph Tanzi). The Autism Genome Project (AGP) work was supported by the following grants: (1) The Hildebrand Foundation (Principal Investigator Joachim F. Hallmayer); (2) Autism Speaks (for the AGP); (3) grants from the National Institutes of Health (NIH) MH61009 (James S. Sutcliffe), MH55135 (Susan E. Folstein), MH55284 (Joseph Piven), HD055782 (Ellen M. Wijsman), NS042165 (Joachim F. Hallmayer); (4) Fundação para a Ciência e Tecnologia (POCTI/39636/ESP/2001) Fundação Calouste Gulbenkian (Astrid Vincente); (5) INSERM, Fondation de France, Fondation Orange, Fondation pour la Recherche Médicale (Catalina Betancur, Marion Leboyer), and the Swedish Science Council (Christopher

Gillberg); (6) The Seaver Foundation (Joseph D. Buxbaum); (7) The Children's Medical & Research Foundation (CMRF), Our Lady's Children's Hospital, Crumlin, Ireland (Sean Ennis); (8) The Medical Research Council (MRC) (Anthony P. Monaco, Anthony J. Bailey). Fresh-frozen brain tissue samples were obtained through the Autism Tissue Program and the Harvard Brain Bank.

Author Contributions L.A.W., D.E.A., M.J.D. and A. Chakravarti led design and execution of joint scan analyses and manuscript writing. Johns Hopkins University–NIMH genome scan team: D.E.A. and A. Chakravarti led study design and analysis of scan; C.W.B. and E.H.C. provided evaluation of phenotype data, phenotype definition from primary data and editing of the manuscript; K. West, A.O.C. and G.H. conducted primary and replication genotyping with allele calling; R.L.T. and A.B.W. performed expression analysis and editing of the manuscript. Autism Consortium–AGRE genome scan team: L.A.W., T.G. and M.J.D. performed data processing and analysis for the genome-wide association scan; S.-C.C., E.M.H., E.M.M., R.S. and S.L.S. provided evaluation of phenotype data and phenotype definition from primary data; S.G., C. Gates, C. Sougnez and C. Stevens led the genotyping team; A.K., J.K., F.K., S.M., B.N. and S.P. performed and evaluated allele calling and advised the analysis; M.J.D., L.A.W., R.T., P.S., S.L.S., J. Gusella and D.A. designed and initiated the study and provided manuscript comments and edits. Replication teams: each replication team provided genotypes, phenotypes and analysis of top ranking SNPs from the combined genome-wide association scan and contributed comments during manuscript preparation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.C. (aravinda@jhmi.edu) or M.J.D. (mjdaly@chgr.mgh.harvard.edu).

The Gene Discovery Project of Johns Hopkins & the Autism Consortium

Writing group Lauren A. Weiss^{1,2†}, Dan E. Arking³, Mark J. Daly^{1,2}, Aravinda Chakravarti³

Johns Hopkins University–NIMH genome scan team Dan E. Arking³, Camille W. Brune⁴, Kristen West³, Ashley O'Connor³, Gina Hilton³, Rebecca L. Tomlinson⁵, Andrew B. West⁵, Edwin H. Cook Jr⁴, Aravinda Chakravarti³

Autism Consortium–AGRE genome scan team Lauren A. Weiss^{1,2†}, Todd Green^{1,2}, Shun-Chiao Chang¹, Stacey Gabriel², Casey Gates², Ellen M. Hanson³, Andrew Kirby^{1,2}, Joshua Korn^{1,2}, Finny Kuruvilla^{1,2}, Steven McCarroll^{1,2}, Eric M. Morrow^{1,2,7,60†}, Benjamin Neale^{1,2}, Shaun Purcell^{1,2}, Rokana Sasanfar^{1,8}, Carrie Sougnez², Christine Stevens², David Altshuler^{1,2}, James Gusella^{1,2}, Susan L. Santangelo¹, Pamela Sklar^{1,2}, Rudolph Tanzi¹, Mark J. Daly^{1,2}

Replication teams: Autism Genome Project Consortium (listed alphabetically)

Richard Anney²⁸, Anthony J. Bailey⁹, Gillian Baird¹⁰, Agatino Battaglia⁵⁶, Tom Berney¹², Catalina Betancur¹³, Sven Bölte¹⁴, Patrick F. Bolton¹⁵, Jessica Brian¹⁶, Susan E. Bryson¹⁷, Joseph D. Buxbaum¹⁸, Ines Cabrito⁵¹, Guiqing Cai¹⁸, Rita M. Cantor¹⁹, Edwin H. Cook Jr⁵, Hilary Coon²⁰, Judith Conroy²⁶, Catarina Correia⁵¹, Christina Corsello²¹, Emily L. Crawford⁴⁶, Michael L. Cuccaro²², Geraldine Dawson⁵⁹, Maretha de Jonge²⁴, Bernie Devlin²⁵, Eftichia Duketis¹⁴, Sean Ennis²⁶, Annette Estes²³, Penny Farrar³⁸, Eric Fombonne²⁷, Christine M. Freitag⁴, Louise Gallagher²⁸, Daniel H. Geschwind²⁹, John Gilbert²², Michael Gill²⁸, Christopher Gillberg⁵³, Jeremy Goldberg³⁰, Andrew Green²⁶, Jonathan Green³¹, Stephen J. Guter⁴, Jonathan L. Haines³², Joachim F. Hallmayer³³, Vanessa Hus²¹, Sabine M. Klauck³⁴, Olena Korvatska⁵⁵, Janine A. Lamb³⁵, Magdalena Laskawiec⁹, Marion Leboyer⁵⁴, Ann Le Couteur¹², Bennett L. Leventhal⁴, Xiao-Qing Liu^{16,44}, Catherine Lord²¹, Linda J. Lotspeich³³, Elena Maestrini⁵⁸, Tiago Magalhaes⁵¹, William Mahoney³⁶, Carine Mantoulan³⁷, Helen McConachie¹², Christopher J. McDougle⁵⁷, William M. McMahon²⁰, Christian R. Marshall⁴⁴, Judith Miller²⁰, Nancy J. Minshew⁴, Anthony P. Monaco³⁸, Jeff Munson²³, John I. Nurnberger Jr⁵⁷, Guiomar Oliveira⁴², Alistair Pagnamenta³⁸, Katerina Papanikolaou³⁹, Jeremy R. Parr⁹, Andrew D. Paterson^{16,44}, Margaret A. Pericak-Vance²², Andrew Pickles⁴⁰, Dalila Pinto⁴⁴, Joseph Piven⁴¹, David J. Posey⁵⁷, Annemarie Poustka^{34†}, Fritz Poustka¹⁴, Regina Regan²⁶, Jennifer Reichert¹⁸, Katy Renshaw⁹, Wendy Roberts¹⁶, Bernadette Roge³⁷, Michael L. Rutter⁴², Jeff Salt⁴, Gerard D. Schellenberg⁴³, Stephen W. Scherer⁴⁴, Val Sheffield⁴⁵, James S. Sutcliffe⁶, Peter Szatmari³⁰, Katherine Tansey²⁸, Ann P. Thompson³⁰, John Tsiatis³⁹, Herman Van Engeland²⁴, Astrid M. Vicente⁵¹, Veronica J. Vieland¹¹, Fred Volkmar⁴⁷, Simon Wallace⁹, Thomas H. Wassink⁴⁸, Ellen M. Wijsman⁴⁹, Kirsty Wing³⁸, Kerstin Wittmeyer³⁷, Brian L. Yaspian⁴⁶, Lonnie Zwaigenbaum⁵⁰; **The Homozygosity Mapping Collaborative for Autism** Eric M. Morrow^{1,2,7,60†}, Seung-Yun Yoo^{2,7,60}, Robert Sean Hill^{2,7,60}, Nahit M. Mukaddes⁶¹, Soher Balkhy⁶², Genaro Gascon^{62,63}, Samira Al-Saad⁶⁵, Asif Hashmi⁶⁴, Janice Ware⁶, Robert M. Joseph⁶⁶, Elaine LeClair⁶, Jennifer N. Partlow^{7,60}, Brenda Barry^{7,60}, Christopher A. Walsh^{2,7,60}; **MGH Oulu study** David Pauls¹, Irma Moilanen⁶⁷, Hanna Ebeling⁶⁷, Marja-Leena Mattila⁶⁷, Sanna Kuusikko⁶⁷, Katja Jussila⁶⁷, Jaakko Ignatius⁶⁷; **MGH Iran study** Rokana Sasanfar^{1,8}, Ala Tolouei⁸, Majid Ghadami⁸, Maryam Rostami⁶⁸, Azam Hosseini⁸, Maryam Valujerdi⁸; **MGH EDSP study** Susan L. Santangelo¹, Kara Andresen⁶⁹, Brian Winkloski¹, Stephen Haddad¹; **Children's Hospital Boston** Lou Kunkel⁷, Zak Kohane⁷, Tram Tran⁷, Sek Won Kong⁷, Stephanie

Brewster O'Neil⁷, Ellen M. Hanson⁶, Rachel Hundley⁶, Ingrid Holm⁷, Heather Peters⁷, Elizabeth Baroni⁷, Aislyn Cangialose⁷, Lindsay Jackson⁷, Lisa Albers⁷, Ronald Becker⁷, Carolyn Bridgemohan⁷, Sandra Friedman⁷, Kerim Munir⁷, Ramzi Nazir⁷, Judith Palfrey⁷, Alison Schonwald⁷, Esau Simmons⁷, Leonard A. Rappaport⁶; **Montreal** Julie Gauthier⁷⁰, Laurent Mottron⁷¹, Ridha Joobar²⁷, Eric Fombonne²⁷, Guy Rouleau⁷¹; **Finland** Karola Rehnstrom^{72,73}, Lennart von Wendt^{72,73} & Leena Peltonen^{72,73,74}

¹Center for Human Genetic Research, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02114, USA. ²Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. ³Center for Complex Disease Genomics, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, Maryland 21205, USA. ⁴Institute for Juvenile Research, Department of Psychiatry, University of Illinois at Chicago, Chicago, Illinois 60612, USA. ⁵Center for Neurodegeneration and Experimental Therapeutics, University of Alabama School of Medicine, Birmingham, Alabama 35294, USA. ⁶Developmental Medicine Center, Children's Hospital Boston, Boston, Massachusetts 02115, USA. ⁷Division of Genetics, Children's Hospital Boston and Harvard Medical School, Boston, Massachusetts 02115, USA. ⁸Special Education Organization, Tehran, Iran. ⁹Department of Psychiatry, University of Oxford, Warneford Hospital, Headington, Oxford OX3 7JX, UK. ¹⁰Newcomen Centre, Guy's Hospital, London SE1 9RT, UK. ¹¹Battelle Center for Mathematical Medicine, The Research Institute at Nationwide Children's Hospital and The Ohio State University, Columbus, Ohio 43205, USA. ¹²Child and Adolescent Mental Health, University of Newcastle, Sir James Spence Institute, Newcastle upon Tyne NE1 4LP, UK. ¹³INSERM U952, Université Pierre et Marie Curie, 75005 Paris, France. ¹⁴Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, J.W. Goethe University Frankfurt, 60528 Frankfurt, Germany. ¹⁵Department of Child and Adolescent Psychiatry, Institute of Psychiatry, London SE5 8AF, UK. ¹⁶Autism Research Unit, The Hospital for Sick Children and Bloorview Kids Rehabilitation, University of Toronto, Toronto, Ontario M5G 1Z8, Canada. ¹⁷Department of Pediatrics and Psychology, Dalhousie University, Halifax, Nova Scotia B3K 6R8, Canada. ¹⁸Laboratory of Molecular Neuropsychiatry, Seaver Autism Center for Research and Treatment, Departments of Psychiatry, Genetics and Genomic Sciences, and Neuroscience, Mount Sinai School of Medicine, New York, New York 10029, USA. ¹⁹Department of Human Genetics, University of California–Los Angeles School of Medicine, Los Angeles, California 90095, USA. ²⁰Psychiatry Department, University of Utah Medical School, Salt Lake City, Utah 84108, USA. ²¹Autism and Communicative Disorders Centre, University of Michigan, Ann Arbor, Michigan 48104, USA. ²²Miami Institute for Human Genomics, University of Miami, Miami, Florida 33136, USA. ²³Departments of Psychology and Psychiatry, University of Washington, Seattle, Washington 98195, USA. ²⁴Department of Child Psychiatry, University Medical Center, Utrecht 3508 GA, The Netherlands. ²⁵University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 19104-6100, USA. ²⁶School of Medicine and Medical Science University College, Dublin 4, Ireland. ²⁷Division of Psychiatry, McGill University, Montreal, Quebec H3A 1A1, Canada. ²⁸Autism Genetics Group, Department of Psychiatry, School of Medicine, Trinity College, Dublin 8, Ireland. ²⁹Department of Neurology, University of California–Los Angeles School of Medicine, Los Angeles, California 90095, USA. ³⁰Department of Psychiatry and Behavioural Neurosciences, McMaster University, Hamilton, Ontario L8N 3Z5, Canada. ³¹Academic Department of Child Psychiatry, Booth Hall of Children's Hospital, Blackley, Manchester M9 7AA, UK. ³²Centre for Human Genetics Research, Vanderbilt University Medical Centre, Nashville, Tennessee 37232, USA. ³³Child and Adolescent Psychiatry and Child Development, Stanford University School of Medicine, Stanford, California 94304, USA. ³⁴Deutsches Krebsforschungszentrum, Molekulare Genomanalyse, 69120 Heidelberg, Germany. ³⁵Centre for Integrated Genomic Medical Research, University of Manchester, Manchester M13 9PT, UK. ³⁶Department of Pediatrics, McMaster University, Hamilton, Ontario L8S 3Z5, Canada. ³⁷Centre d'Etudes et de Recherches en Psychopathologie, Université de Toulouse Le Miral, Toulouse 31058, France. ³⁸Wellcome Trust Centre for Human Genetics, University of Oxford, OX3 7BN, UK. ³⁹University Department of Child Psychiatry, Athens University, Medical School, Agia Sophia Children's Hospital, Athens 115, Greece. ⁴⁰Department of Medicine, School of Epidemiology and Health Science, University of Manchester, Manchester M13 9PT, UK. ⁴¹Carolina Institute for Developmental Disabilities, University of North Carolina, Chapel Hill, North Carolina 27599-3366, USA. ⁴²Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, London SE5 8AF, UK. ⁴³Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19108, USA. ⁴⁴The Centre for Applied Genomics and Program in Genetics and Genome Biology, The Hospital for Sick Children and Department of Molecular Genetics, University of Toronto, Toronto, Ontario M5G 1L7, Canada. ⁴⁵Department of Pediatrics and Howard Hughes Medical Institute Carver College of Medicine, University of Iowa, Iowa City, Iowa 52242, USA. ⁴⁶Vanderbilt Kennedy Center and Center for Molecular Neuroscience, Vanderbilt University, Nashville, Tennessee 37232, USA. ⁴⁷Child Study Centre, Yale University, New Haven, Connecticut 06510, USA. ⁴⁸Department of Psychiatry, Carver College of Medicine, Iowa City, Iowa 52242, USA. ⁴⁹Department of Biostatistics and Medicine, University of Washington, Seattle, Washington 98195, USA. ⁵⁰Department of Pediatrics, University of Alberta, Edmonton, Alberta T6G 2J3, Canada. ⁵¹Instituto Nacional de Saude Dr Ricardo Jorge Instituto Gulbenkian de Ciencia Lisbon, 1600-560 Portugal. ⁵²Hospital Pediatrico de Coimbra, Coimbra, 3000-300 Portugal. ⁵³Department of Child and Adolescent Psychiatry, Göteborg University, Göteborg S41345, Sweden. ⁵⁴INSERM U995, Department of Psychiatry, Groupe hospitalier Henri Mondor–Albert Chenevier, AP-HP, Créteil, France. ⁵⁵Department of Medicine, University of Washington, Seattle, Washington 98195, USA. ⁵⁶Stella Maris Institute, Department of Child and Adolescent Neurosciences, 56018 Calambrone (Pisa), Italy. ⁵⁷Department of Psychiatry, Indiana

University School of Medicine, Indianapolis 46202, USA. ⁵⁸Department of Biology, University of Bologna, 40126 Bologna, Italy. ⁵⁹Autism Speaks, New York, New York 10016, USA ⁶⁰Department of Neurology and Howard Hughes Medical Institute, Beth Israel Deaconess Medical Center, Boston, Massachusetts 02215, USA. ⁶¹Department of Child Psychiatry, Istanbul Faculty of Medicine, Istanbul University, 34452 Istanbul, Turkey. ⁶²Department of Neurosciences and Pediatrics, King Faisal Specialist Hospital and Research Centre, Jeddah 11211, Kingdom of Saudi Arabia. ⁶³Clinical Neurosciences & Pediatrics, Brown University School of Medicine, Providence, Rhode Island, USA. ⁶⁴Department of Neurology, Combined Military Hospital, Lahore, Pakistan. ⁶⁵Kuwait Center for Autism, Kuwait City 73455, Kuwait. ⁶⁶Department of Anatomy and Neurobiology, Boston University School of Medicine, Boston, Massachusetts 02118, USA. ⁶⁷Department of Child Psychiatry and Department of Clinical Genetics, Oulu

University Hospital and Oulu University, Oulu FIN-90029, Finland. ⁶⁸Medical Genetic Diagnosis Department, National Institute for Genetic Engineering and Biotechnology, Tehran, Iran. ⁶⁹Casa de Corazon, Taos, New Mexico 87571, USA. ⁷⁰Centre de recherche du CHUM, Hôpital Notre-Dame, Montréal, H2L 4M1 Quebec, Canada. ⁷¹Sainte-Justine Hospital Research Center, Université de Montréal, Montréal, H3T 1C5 Quebec, Canada. ⁷²Department of Medical Genetics, University of Helsinki, Helsinki, FIN-00014, Finland. ⁷³Department of Molecular Medicine, National Public Health Institute, Helsinki, FIN-00014, Finland. ⁷⁴Wellcome Trust Sanger Institute, Hinxton CB10 1SA, UK. †Present addresses: Department of Psychiatry, Institute for Human Genetics, Center for Neurobiology and Psychiatry, UCSF, San Francisco, California, USA (L.A.W.); Department of Molecular Biology, Cell Biology and Biochemistry, and Institute for Brain Science, Brown University, Providence, Rhode Island, USA (E.M.M.). ‡Deceased.

METHODS

All samples used in this study arose from investigations approved by the individual and respective Institutional Review Boards in the USA and at international sites where relevant. Informed consent was obtained for all adult study participants; for children under age 18, both the consent of the parents or guardians and the assent of the child were obtained.

Primary study samples: AGRE samples. The Autism Genetic Resource Exchange (AGRE) curates a collection of DNA and phenotypic data from multiplex families with autism spectrum disorder (ASD) available for genetic research³. We genotyped individuals from 801 families, selecting those with at least one child meeting criteria for autism by the Autism Diagnostic Interview-Revised (ADI-R)²⁵, whereas the second affected child had an AGRE classification of autism, broad spectrum (patterns of impairment along the spectrum of pervasive developmental disorders, including pervasive developmental disorder not otherwise specified (PDD-NOS) and Asperger's syndrome) or not quite autism (NQA, individuals who are no more than one point away from meeting autism criteria on any or all of the social, communication, and/or behaviour domains and meet criteria for 'age of onset'; or, individuals who meet criteria on all domains, but do not meet criteria for the 'age of onset'). We excluded probands with widely discrepant classifications of affection status via the ADI-R and Autism Diagnostic Observation Schedule (ADOS) that could not be reconciled. We also excluded families with known chromosomal abnormalities (where karyotyping was available), and those with inconsistencies in genetic data (generating excess Mendelian segregation errors or showing genotyping failure on a test panel of 24 SNPs used to check gender and sample identity with the full array data). The self-reported race/ethnicity of these samples is 69% white, 12% Hispanic/Latino, 10% unknown, 5% mixed, 2.5% each Asian and African American, less than 1% native Hawaiian/Pacific Islander and American Indian/native Alaskan.

Primary study samples: NIMH samples. The NIMH Autism Genetics Initiative maintains a collection of DNA from multiplex and simplex families with ASD. We genotyped individuals from 341 nuclear families, 258 of which were independent of the AGRE data set, with at least one child meeting criteria for autism by the ADI-R, and a second child considered affected using the same criteria as described for the AGRE data set above. Similar exclusion criteria were used, including known chromosomal abnormalities and excess non-Mendelian inheritance. The self-reported race/ethnicity of these samples is 83% white, 4% Hispanic, 2% unknown, 7% mixed, 3% Asian and 1% African American.

Primary study samples: merged data set for primary screening. We used the Birdseed algorithm for genotype calling at both genotyping centres^{26,27}. As 324 individuals were genotyped at both centres, we performed a concordance check. One sample showed substantial differences between the two centres, but no excess of Mendelian errors, indicating that a sample mix-up occurred in which each centre genotyped a different sibling that was identified as the same sample. Excluding this sample, overall genotype concordance between the two centres was 99.72%.

Before merging data, we examined the distribution of chi-squared values and used a series of quality control (QC) filters designed to identify a robust set of SNPs. We discovered that filtering AGRE genotypes to 98% completeness and less than 10 Mendelian errors (MEs) was sufficient to remove SNPs that artificially inflated the chi-squared distribution for SNPs with MAF > 0.05. For MAF < 0.05, we observed much greater inflation ($\lambda = 1.17$), due entirely to a strong excess of SNPs with under-transmission of the minor allele (OR < 1). Whereas the same filters yielded high-quality results for SNPs with over-transmission of the minor allele ($\lambda = 1.04$), we found that much stricter filtering was required for rarer SNPs with OR < 1 (missing data < 0.005). This is not unexpected based on a well-documented bias in the TDT: if missing data are preferentially biased against heterozygotes or rare homozygotes, significant, artificial over-transmission of the common allele is expected^{28,29}. To achieve comparable quality for the NIMH data set, we filtered on 96% completeness and fewer than 4 MEs. Our final QQ plot for the combined data set is shown in Supplementary Fig. 1 and has a $\lambda \approx 1.03$, less than that observed in the Wellcome Trust Case Control Consortium paper for five of the seven phenotypes studied³⁰. The combined data set, consisting of 1,031 families (856 with two parents) and a total of 1,553 affected offspring, was used for association testing.

For linkage analyses, the combined AGRE/NIMH data set was further merged with Illumina 550K genotype data generated at the Children's Hospital of Philadelphia (CHOP) and available from AGRE, adding ~300 nuclear families (1,499 samples). We used the extensive overlap of samples between the AGRE/NIMH and the CHOP data sets (2,282 samples) to select an extremely high quality set of SNPs for linkage analysis. Specifically, we required SNPs to be on both the Affymetrix 500K/5.0 and Illumina 550K platforms, with >99.5% concordance across platforms. We further restricted SNPs to MAF > 0.2, <1%

missing data, Hardy-Weinberg $P > 0.01$, and no more than 1 ME. This left ~36,000 SNPs of outstanding quality. For autosomal SNPs, we further pruned using PLINK to remove SNPs with $r^2 > 0.1$, yielding 16,311 SNPs.

Replication samples: NIMH control samples. Controls obtained from the NIMH Genetics Repository were genotyped on the Affymetrix 500K platform at the Broad Institute Genetic Analysis Platform for another study⁶. Of these, 1,494 matched well with our sample, and were used as controls to compare with the cases and parents in our study.

Replication samples: Montreal samples. Subjects diagnosed with autism spectrum disorders with both of their parents were recruited from clinics specializing in the diagnosis of Pervasive Developmental Disorders (PDD), readaptation centres, and specialized schools in the Montreal and Quebec City regions, Canada, as described³¹. Subjects with ASD were diagnosed by child psychiatrists and psychologists expert in the evaluation of ASD. Evaluation based on the Diagnostic and Statistical Manual of Mental Disorders (DSM) criteria included the use of the ADI-R²⁵ and the ADOS³². As an additional screening tool for the diagnosis of ASD, the Autism Screening Questionnaire, which is derived from the ADI-R, was completed³³. Furthermore, all proband medical charts were reviewed by a child psychiatrist expert in PDD to confirm their diagnosis and exclude subjects with any co-morbid disorders. Exclusion criteria were: (1) an estimated mental age <18 months; (2) a diagnosis of Rett syndrome or childhood disintegrative disorder; and (3) evidence of any psychiatric and neurological conditions including: birth anoxia, rubella during pregnancy, fragile X syndrome, encephalitis, phenylketonuria, tuberous sclerosis, Tourette and West syndromes. Subjects with these conditions were excluded based on parental interview and chart review. However, participants with a co-occurring diagnosis of semantic-pragmatic disorder (owing to its large overlap with PDD), attention deficit hyperactivity disorder (seen in a large number of patients with ASD during development), and idiopathic epilepsy (related to the core syndrome of ASD) were eligible for the study.

Replication samples: Santangelo EDSP family samples. Families were ascertained for having one or more autistic children and at least one non-autistic child aged 16 or older for an extremely discordant sibling-pair linkage study. Recruitment took place in Massachusetts and surrounding states through contacts with parent support and patient advocacy groups, brochures, newsletters and the study website. Parents were interviewed about their children, and non-autistic children were interviewed about themselves. An informant/caregiver, usually the proband's mother, was interviewed using the ADI-R to confirm the diagnosis of autism at age 4–5 years^{25,34}. Families were included if the affected children met Diagnostic and Statistical Manual of Mental Disorders-IV (DSM-IV) criteria for autistic disorder and their non-autistic siblings (aged 16 and older) did not display any of the broader autism phenotype traits, which were assessed with the (M-PAS-R), the Pragmatic Language Scale (PLS), and the Friendship Interview^{35,36}. Probands were excluded if they had medical conditions associated with autism such as fragile X syndrome or gross CNS injury, or if they were under 4 years of age, owing to the possible uncertainty in diagnosis at younger ages. Twenty-nine families met eligibility criteria for the study and comprised the final sample for analysis.

Replication samples: high functioning autism family samples. Families were included if their affected child had been previously diagnosed with Autism or Asperger's syndrome, had a level of intellectual functioning above the range of mental retardation (that is, full scale, verbal and performance IQ > 70), chronological age between 6 and 21 years, and an absence of significant medical or neurological disorders (including fragile X syndrome and tuberous sclerosis). Families were ascertained and recruited through the Acute Residential Treatment (ART) programmes and outpatient child and adolescent services at McLean Hospital, as well as through associated hospitals and clinics. Brochures and a website were also used. Thirty-three families (133 participants) were enrolled in the study. Participation was voluntary.

Replication samples: MGH-Finnish collaborative samples. Altogether 58 individuals with a diagnosis of high functioning autism (HFA) or Asperger's syndrome were recruited in Finland. Fifty-two children and adolescents aged 8–15 years were identified from patient records at the Oulu University Hospital in 2003. These children and adolescents have been evaluated for HFA/Asperger's syndrome at the Oulu University Hospital. In addition, six children (3 boys, 3 girls) 11 years of age were recruited from an epidemiological study conducted in 2001 (ref. 37).

All participants had full-scale IQ scores greater than or equal to 80 measured with the Wechsler Intelligence Scale for Children—Third Revision³⁸. Furthermore, none of the children subjects was diagnosed with other developmental disorders (for example, dysphasia, fragile X syndrome). Clinical diagnoses of HFA/Asperger's syndrome were confirmed by administering the ADI-R²⁵ and the ADOS³². Of the 58 participants with HFA/Asperger's syndrome, 35 met the diagnostic criteria for Asperger's syndrome and 21 met the diagnostic criteria for HFA

according to ICD-10 (International Classification of Diseases v. 10) diagnostic criteria³⁹. Two participants met diagnostic criteria for PDD-NOS; these participants were excluded owing to their manifesting different and less severe symptoms than our sample of children with HFA or Asperger's syndrome.

Replication samples: Children's Hospital Boston samples. Probands with a documented history of clinical diagnosis of ASD were recruited at Children's Hospital Boston. To participate, they had to be over 24 months of age and have at least one biological parent or an affected sibling available. Subjects were excluded if they had an underlying metabolic disorder or any chronic systemic disease, an acquired developmental disability (for example, birth asphyxia, trauma-related injury, meningitis, etc.), or cerebral palsy. All participants provided informed consent and a phenotyping battery was performed including the ADOS, the ADI-R and other measures to assess cognitive status. Seventy-five per cent of subjects with a clinical diagnosis met strict research criteria for ASD on both ADI-R and ADOS. In addition, a complete family and medical history was obtained.

Replication samples: homozygosity mapping collaborative for autism (HMCA) samples. Families with cousin marriages and children affected by ASD with or without mental retardation were recruited by multiple collaborators in the HMCA. The patients from Istanbul were evaluated by a child psychiatrist (N. M. Mukaddes) trained in the ADOS and ADI-R, and who made diagnoses according to DSM-IV-TR criteria and the Childhood Autism Rating Scale (CARS). Patients from Kuwait were enrolled from the Kuwait Centre for Autism by S. Al-Saad. In Jeddah, Saudi Arabia, patients were evaluated by both a developmental paediatrician (S. Balkhy) and a paediatric neurologist (G. Gascon) and diagnoses were based on DSM-IV-TR criteria. In Lahore, Pakistan, a neurologist (A. Hashmi) with training in the ADOS and ADI-R diagnosed patients using DSM-IV-TR criteria. In most settings, patients were enrolled from tertiary clinical centres and these patients had standard of care neuromedical assessments, including physical examination, medical and neurological history, fragile X testing, and other genetic and metabolic testing when indicated. MRI was obtained for patients in whom a brain malformation was suspected or seizures were present. In addition, IQ scores (usually from the Stanford-Binet) and adaptive behaviour measures were obtained from the patients' existing medical records. Secondary assessments were conducted on the most informative pedigrees by the Boston clinical team in collaboration with local multi-disciplinary teams. Clinical members of the Boston team included: developmental psychologists (J. Ware, E. LeClaire, R. M. Joseph), paediatric neurologists (G. H. Mochida, A. Poduri), a clinical geneticist (W.-H. Tan) and a neuropsychiatrist (E. M. Morrow). The secondary assessment battery was designed to obtain a comprehensive description of current and historical autism symptomatology, cognitive and adaptive functioning, and neurological and physical morphological status in the patient and pedigree. The secondary assessment included: neurological examination; genetic dysmorphology examination; the CARS; the Social Communication Questionnaire administered with probing on par with the ADI-R by ADI-R reliable examiners; the ADOS (usually module 1); the Vineland Adaptive Behaviour Scales, second edition (VABS-II); Kaufman Brief Intelligence Test, second edition (KBIT-II). ADOS assessments were videotaped and dysmorphology findings were photographed for archival purposes.

Replication samples: AGP samples. Individuals typically received at least two of three evaluations for autism symptoms: ADI-R, ADOS and clinical evaluation. Of the 1,679 affected individuals from 1,443 families, 966 met criteria for autism on the ADI-R and ADOS and most of these also had a clinical evaluation of autism; 160 affected individuals met criteria for autism on one of the two diagnostic instruments (ADI-R, ADOS) but were missing information on the other instrument; and, 553 individuals met criteria for spectrum disorder on one or both instruments. Affected individuals were recruited from both simplex and multiplex families, 71% of this sample being from multiplex families. Most of the families were of European ancestry (83%).

Replication samples: Finnish autism family samples. Families were recruited through university and central hospitals. Detailed clinical and medical examinations were performed by experienced child neurologists as described elsewhere⁴⁰. Diagnoses were based on ICD-10³⁹ and DSM-IV⁴¹ diagnostic nomenclatures. Families with known associated medical conditions or chromosomal abnormalities were excluded from the study. A total of 106 families included 400 individuals for whom genotype data was available. Of these, 111 had a diagnosis of infantile autism and 13 a diagnosis of Asperger's syndrome. All families were Finnish, except for one family where the father was Turkish.

Replication samples: Iranian trio samples. Eligible participants in this study were Iranian families with at least one child affected with ASD, including cases of autistic disorder, Asperger's syndrome and PDD-NOS. Eighty families (282 individuals) from Iran were ascertained and assessed. This sample was ascertained by screening and diagnostic testing of over 90,000 preschool children from Tehran in 2004. Diagnoses of children were made according to DSM-IV criteria via the ADI-R and the ADOS. Patients with abnormal karyotypes and

dysmorphic features were excluded. Most of the families were father-mother-child trios but some had more than one affected child. All affected biological siblings were assessed with the same diagnostic tools. We have ascertained and assessed 80 families (282 individuals) from Iran.

Affymetrix genotyping. The AGRE samples were genotyped on Affymetrix 5.0 chips at the Genetic Analysis Platform of the Broad Institute, using standard protocols. The 5.0 chip was designed to genotype nearly 500,000 SNPs across the genome to enable genome-wide association studies^{26,27}. The NIMH controls were genotyped at the Broad Institute using the Affymetrix 500K Sty and Nsp chips, using a similar protocol⁶. The Autism Consortium and Montreal replication samples were also genotyped at the Broad Institute under the same conditions. The NIMH autism samples were genotyped at the Johns Hopkins Center for Complex Disease on the Affymetrix 500K (Nsp and Sty) and 5.0 platforms using similar standard protocols.

Genotype calling for the 5.0 arrays was performed by Birdseed^{26,27} and for the 500K arrays was performed by BRLMM. As basic QC filters for the data generated at the Broad Institute, we required that genotyping was >95% complete for each individual, and that each family had fewer than 10,000 Mendelian inheritance errors across the genome. We also required that each SNP had >95% genotyping, fewer than 15 Mendelian errors, Hardy-Weinberg equilibrium $P > 10^{-10}$, and minor allele frequency greater than 1%. For the AGRE sample, this left 2,883 high-quality individuals genotyped for 399,147 SNPs with 99.6% average call rate. The basic filters for the data generated at Johns Hopkins were individual call rates >95% for 5.0 arrays and >90% for 500K arrays data, fewer than 5,000 Mendelian errors per family. Only monomorphic SNPs and those with greater than 50% missing data were dropped, for 498,216 SNPs. Our combined data set had nearly 365,000 SNPs passing QC.

Sequenom genotyping. SNPs were assayed using Sequenom technology for the AGP samples at three centres, namely Gulbenkian, Mt Sinai and Oxford: DNA from 1,629 families representing numerous recruiting sites was genotyped for 54 SNPs. SNPs with >3% missing data, namely rs4690464, rs10513025, and rs17088296, were excluded from analysis. The next step in our QC process was to remove families with ≥ 4 Mendelian errors, out of 51 remaining loci, under the assumption that this indicated pedigree errors. Data from 110 families were removed owing to Mendelian errors. Thereafter, SNPs were removed if they showed excessive Mendelian errors (>16) in the remaining families. Using this criterion, two more SNPs, rs155437 and rs1925058, were removed from analysis. It was apparent that DNA quality varied by study site and could be responsible for concomitant genotype quality differences. Therefore, we also evaluated rate of missing genotypes per locus and study site. Our analyses showed that DNA from a few population samples showed excess missingness for two SNPs, rs4742408 and rs7869239, relative to the remaining population samples. Specifically three population samples showed more than 7% missing genotypes for rs4742408 and rs7869239 whereas the remaining population samples had about 1% or less missing genotypes. Therefore, for these loci we deleted genotypes only from the samples showing excess missingness. As a final QC step, we then evaluated missing genotypes for the remaining loci. If more than five loci were missing genotypes, the individual's data was removed from analysis. By this criterion 76 additional families became uninformative for family-based association analysis, leaving 1,443 families for association analysis. The Finnish autism samples were genotyped in the Peltonen laboratory, and the Iranian trios were genotyped at the Broad Institute using very similar protocols. All samples were genotyped using aliquots from the same pooled primers and probes.

Copy number analysis. Because of previous reports of two large (>1 Mb), independent *de novo* deletions spanning this locus⁴², we assessed the region surrounding rs10513025 and the entire *SEMA5A* locus for copy number variation that could either explain or provide independent evidence of the importance of this region to autism using Birdsuite²⁶ to analyse all Affymetrix 5.0 samples. Birdsuite genotypes previously annotated common copy number polymorphisms²⁷ and in parallel searches for novel copy number variants (CNVs) using an HMM. Probe coverage in the region was good, with no 50-kb window having fewer than 10 probes and an average spacing between probes of 2.5 kb, allowing very good sensitivity for CNVs greater than 25 kb. We found no deletions or duplications near this SNP, nor any overlapping the gene *SEMA5A*. The closest CNS upstream and downstream of this SNP appeared to be a rare (~2–3% frequency, previously annotated CNP) 40-kb deletion from 288 kb from the 3' end of *SEMA5A*, and a rare (~1% frequency, novel) 20-kb deletion 356 kb upstream of the 5' end of *SEMA5A*. Each of these appeared to be segregating polymorphisms, but far outside of the boundaries of *SEMA5A* and *TAS2R1* and far beyond the linkage disequilibrium block containing rs10513025.

Expression analysis. Fresh-frozen brain tissue samples dissected from the cortex (Brodman area 19) were obtained through the Autism Tissue Program (<http://www.atportal.org>) from the Harvard Brain Bank and the NICHD Brain and Tissue Bank at the University of Maryland from 20 samples with a primary

diagnosis of autism, and 10 controls. Total RNA was extracted using TRIzol reagent (Invitrogen) according to the manufacturer's protocol. Complementary DNA (cDNA) was generated from 8 µg of total RNA using the Superscript III First-Strand Synthesis kit (Invitrogen). cDNA was diluted 1:5 in 10 mM Tris and 1 µl of diluted cDNA was used per 10 µl PCR reaction. Quantitative real-time PCR was performed on a Lightcycler 480 (Roche Applied Science) using 2× Taqman Gene Expression Master Mix and probes obtained from Applied Biosystems (ABI): *SEMA5A* (Hs01549381_m1), *MAP2* (Hs01103234_g1), *TBP* (Hs00920497_m1), *GAPDH* (4333764F). For multiplex reactions, 0.5 µl FAM-labelled *SEMA5A* probe and 0.5 µl VIC-labelled *MAP2* probe were used per 10 µl reaction. The amount of *SEMA5A* relative to *MAP2* was determined for each case using the $\Delta\Delta C_t$ method⁴³. Comparison of *SEMA5A* to *TBP* and *GAPDH* yielded similar results. Logistic regression was performed on autism status, adjusting for age at death, post-mortem interval, sex and *SEMA5A* expression, with a 1-sided *P*-value reported for the association of lower *SEMA5A* expression with autism status.

Determination of significance. To determine an appropriate experimental threshold for genome-wide significance, permutation was performed on this data set by gene-dropping, and genome-wide significance was estimated by taking the lowest *P*-value from each of 1,000 permuted data sets and using the 50th as a threshold for $P < 0.05$ experiment-wide significance ($P < 2.5 \times 10^{-7}$). To calculate an estimate of the effective number of tests (T_{eff}), we used the following algorithm: (1) start with the most 5' SNP on a chromosome ($\text{SNP}_{i,j}$), where i = chromosome and j = SNP position, and calculate pairwise LD with all downstream SNPs within 1 Mb ($r^2[\text{SNP}_{1,1} \times \text{SNP}_{1,n}]$). (2) For $\text{SNP}_{1,1}$, $T_{\text{eff}(1,1)} = 1 - \max(r^2[\text{SNP}_{1,1} \times \text{SNP}_{1,n}])$. (3)

For chromosome i , $T_{\text{eff}(i)} = \sum_{j=1}^m T_{\text{eff}(i,j)}$, where m = the total number of SNPs on

a chromosome. (4) $T_{\text{eff}} = \sum_{i=1}^{23} T_{\text{eff}(i)}$. Because this algorithm only accounts for pairwise LD, it provides a conservative estimate of the number of effective tests.

25. Lord, C., Rutter, M. & Le Couteur, A. Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J. Autism Dev. Disord.* **24**, 659–685 (1994).
26. Korn, J. M. *et al.* Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nature Genet.* **40**, 1253–1260 (2008).
27. McCarroll, S. A. *et al.* Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nature Genet.* **40**, 1166–1174 (2008).
28. Hirschhorn, J. N. & Daly, M. J. Genome-wide association studies for common diseases and complex traits. *Nature Rev. Genet.* **6**, 95–108 (2005).
29. Mitchell, A. A., Cutler, D. J. & Chakravarti, A. Undetected genotyping errors cause apparent overtransmission of common alleles in the transmission/disequilibrium test. *Am. J. Hum. Genet.* **72**, 598–610 (2003).
30. The Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).
31. Gauthier, J. *et al.* Autism spectrum disorders associated with X chromosome markers in French-Canadian males. *Mol. Psychiatry* **11**, 206–213 (2006).
32. Lord, C. *et al.* The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. *J. Autism Dev. Disord.* **30**, 205–223 (2000).
33. Berument, S. K., Rutter, M., Lord, C., Pickles, A. & Bailey, A. Autism screening questionnaire: diagnostic validity. *Br. J. Psychiatry* **175**, 444–451 (1999).
34. Le Couteur, A. *et al.* Autism Diagnostic Interview: A standardized investigator-based instrument. *J. Autism Dev. Disord.* **19**, 363–387 (1989).
35. Tyrer, P. J. *Personality Disorders: Diagnosis, Management, and Course* (Wright, 1988).
36. Landa, R. *et al.* Social language use in parents of autistic individuals. *Psychol. Med.* **22**, 245–254 (1992).
37. Mattila, M. L. *et al.* An epidemiological and diagnostic study of Asperger syndrome according to four sets of diagnostic criteria. *J. Am. Acad. Child Adolesc. Psychiatry* **46**, 636–646 (2007).
38. Wechsler, D. *Wechsler Intelligence Scale for Children* Third edn (The Psychological Corporation, 1991).
39. World Health Organization. *The ICD-10 Classification of Mental and Behavioural Disorders. Diagnostic Criteria for Research* (WHO, 1993).
40. Auranen, M. *et al.* A genomewide screen for autism-spectrum disorders: evidence for a major susceptibility locus on chromosome 3q25–27. *Am. J. Hum. Genet.* **71**, 777–790 (2002).
41. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV)* 4 edn (APA, 1994).
42. Marshall, C. R. *et al.* Structural variation of chromosomes in autism spectrum disorder. *Am. J. Hum. Genet.* **82**, 477–488 (2008).
43. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_t}$ method. *Methods* **25**, 402–408 (2001).

Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution

Sohrab P. Shah^{1,2*}, Ryan D. Morin^{3*}, Jaswinder Khattra¹, Leah Prentice¹, Trevor Pugh³, Angela Burleigh¹, Allen Delaney³, Karen Gelmon⁴, Ryan Guliany¹, Janine Senz², Christian Steidl^{2,5}, Robert A. Holt³, Steven Jones³, Mark Sun¹, Gillian Leung¹, Richard Moore³, Tessa Severson³, Greg A. Taylor³, Andrew E. Teschendorff⁶, Kane Tse¹, Gulisa Turashvili¹, Richard Varhol³, René L. Warren³, Peter Watson⁷, Yongjun Zhao³, Carlos Caldas⁶, David Huntsman^{2,5}, Martin Hirst³, Marco A. Marra³ & Samuel Aparicio^{1,2,5}

Recent advances in next generation sequencing^{1–4} have made it possible to precisely characterize all somatic coding mutations that occur during the development and progression of individual cancers. Here we used these approaches to sequence the genomes (>43-fold coverage) and transcriptomes of an oestrogen-receptor- α -positive metastatic lobular breast cancer at depth. We found 32 somatic non-synonymous coding mutations present in the metastasis, and measured the frequency of these somatic mutations in DNA from the primary tumour of the same patient, which arose 9 years earlier. Five of the 32 mutations (in *ABCB11*, *HAUS3*, *SLC24A4*, *SNX4* and *PALB2*) were prevalent in the DNA of the primary tumour removed at diagnosis 9 years earlier, six (in *KIF1C*, *USP28*, *MYH8*, *MORC1*, *KIAA1468* and *RNASEH2A*) were present at lower frequencies (1–13%), 19 were not detected in the primary tumour, and two were undetermined. The combined analysis of genome and transcriptome data revealed two new RNA-editing events that recode the amino acid sequence of *SRP9* and *COG3*. Taken together, our data show that single nucleotide mutational heterogeneity can be a property of low or intermediate grade primary breast cancers and that significant evolution can occur with disease progression.

Lobular breast cancer is an oestrogen-receptor-positive (ER⁺, also known as ESRI⁺) subtype of breast cancer (approximately 15% of all breast cancers). It is usually of low-intermediate histological grade and can recur many years after initial diagnosis. To interrogate the genomic landscape of this class of tumour, we re-sequenced^{1–4} the DNA from a metastatic lobular breast cancer specimen (89% tumour cellularity; Supplementary Fig. 1) at approximately 43.1-fold aligned, haploid reference genome coverage (120.7 gigabases (Gb) aligned paired-end sequence; Supplementary Fig. 2, Table 1 and Supplementary Methods). Deep high-throughput transcriptome sequencing (RNA-seq)⁵ performed on the same sample generated 160.9-million reads that could be aligned (Supplementary Table 1, see also Supplementary Fig. 2 and Supplementary Methods). The saturation of the genome (Table 1) and RNA-seq (Supplementary Table 1) libraries for single nucleotide variant (SNV) detection is discussed in Supplementary Information. The aligned (hg18) reads were used to identify (Supplementary Fig. 2) the presence of genomic aberrations, including SNVs (Supplementary Table 2), insertions/deletions (indels), gene fusions, translocations, inversions and copy number alterations (Supplementary Methods). We examined predicted

coding indels and predicted inversions (coding or non-coding; Supplementary Methods); however, all of the events that were validated by Sanger re-sequencing were also present in the germ line (Supplementary Tables 3 and 4). None of the 12 predicted gene fusions revalidated. We also computed the segmental copy number (Supplementary Methods and Supplementary Table 5a) from aligned reads, and revalidated high level amplicons by fluorescence *in situ* hybridization (FISH) (Supplementary Table 5b), revealing the presence of a new low-level amplicon in the *INSR* locus (Supplementary Fig. 3).

We identified coding SNVs from aligned reads, using a Binomial mixture model, SNVMix (Supplementary Table 2, Methods and Supplementary Appendix 1). From the RNA-seq (WTSS-PE) and genome (WGSS-PE) libraries we predicted 1,456 new coding non-synonymous SNVMix variants (Supplementary Table 2). After the removal of pseudogene and HLA sequences (1,178 positions remaining) and after primer design, we re-sequenced (Sanger amplicons) 1,120 non-synonymous coding SNV positions in the tumour DNA and normal lymphocyte DNA. Some 437 positions (268 unique to WGSS-PE, 15 unique to WTSS-PE, and 154 in common) were confirmed as non-synonymous coding variants. Of these, 405 were new

Table 1 | Summary of sequence library coverage

	WGSS-PE	WTSS-PE
Total number of reads	2,922,713,774	182,532,650
Total nucleotides (Gb)	140.991	7.108
Number of aligned reads	2,502,465,226	160,919,484
Aligned nucleotides (Gb)	120.718	6.266
Estimated error rate	0.021	0.013
Estimated depth (non-gap regions)	43.114	NA
Canonically aligned reads	2,294,067,534	109,093,616
Exons covered	93.5 at >10 reads; 95.7 at >5 reads	82,200 at 10 reads (see also Supplementary Table 1)
Reads aligned canonically (%)	78.49	67.79
Unaligned reads	420,248,548	21,613,166
Mean read length (bp)	48.24	38.94

The WGSS-PE column shows the genome paired-end read coverage for DNA from the metastatic pleural effusion sample. The WTSS-PE column shows coverage for the complementary DNA reads from the matched transcriptome libraries of the metastatic pleural effusion. Coverage of exon bases in the reference genome (hg18) is shown at 5 or more reads per position, and 10 or more reads per position for the metastatic genome. bp, base pairs; NA, not applicable.

¹Molecular Oncology, ²Centre for Translational and Applied Genomics, ³Michael Smith Genome Sciences Centre, BC Cancer Agency, 675 West 10th Avenue, Vancouver V5Z 1L3, Canada. ⁴Medical Oncology, BC Cancer Agency, 600 West 10th Avenue, Vancouver V5Z 1L3, Canada. ⁵Department of Pathology, University of British Columbia, G227-2211 Wesbrook Mall, British Columbia, Vancouver V6T 2B5, Canada. ⁶Cancer Research UK, Cambridge Research Institute, Li Ka Shing Centre, Robinson Way, Cambridge CB2 0RE, UK. ⁷Deeley Research Centre, BC Cancer Agency, Victoria V8R 6V5, Canada.

*These authors contributed equally to this work.

Table 2 | Somatic coding sequence SNVs validated by Sanger sequencing

Gene	Description	Position	Source	Allele change	Amino acid change	Protein domain affected	Expression (sequenced bases per exonic base)	Allelic expression bias (R, NR allele)	Copy number classification (HMM state)
<i>ABCB11</i>	Bile salt export pump (ATP-binding cassette sub-family B member 11)	2:169497197	WGSS	C>T	R>H	Transmembrane helix 3	0.3	1, 1	Amplification (4)
<i>HAUS3</i>	HAUS3 coiled-coil protein (C4orf15)	4:2203607	WGSS, WTSS	C>T	V>M	Unknown	14.1	4, 23	Neutral (2)
<i>CDC6</i>	Cell division control protein 6 homologue	17:35701114	WGSS, WTSS	G>A	E>K	N-terminal, unknown	2.7	3, 3	Amplification (4)
<i>CHD3</i>	Chromodomain-helicase-DNA-binding protein 3	17:7751231	WGSS	G>A	E>K	Unknown, C-terminal	3.9	41, 11 (Q < 0.01)	Neutral (2)
<i>DLG4</i>	Disks large homologue 4	17:7052251	WGSS	G>A	P>L	Unknown, N-terminal	5.5	7, 1	Neutral (2)
<i>ERBB2</i>	Receptor tyrosine-protein kinase erb-b2	17:35133783	WGSS, WTSS	C>G	I>M	Kinase domain	67.1	62, 35	Amplification (4)
<i>FGA</i>	Fibrinogen alpha chain	4:155726802	WGSS	C>T	W>stop	Fibrinogen a/b/c domain	0.01	NA	Gain (3)
<i>GOLGA4*</i>	Golgin subfamily A member 4	3:37267947	WGSS, WTSS	G>C	E>Q	Unknown, N-terminal	111.8	37, 12	Gain (3)
<i>GSTCD</i>	Glutathione S-transferase C-terminal domain-containing protein	4:106982671	WGSS, WTSS	G>C	E>Q	Unknown, C-terminal	23.2	23, 8	Neutral (2)
<i>KIAA1468*</i>	LisH domain and HEAT repeat-containing protein	18:58076768	WGSS, WTSS	G>C	R>T	ARM type fold	36.1	23, 11	Neutral (2)
<i>KIF1C</i>	Kinesin-like protein KIF1C	17:4848025	WGSS, WTSS	G>C	K>N	Kinesin motor domain	28.5	16, 13	Neutral (2)
<i>KLHL4</i>	Kelch-like protein 4	X:86659878	WGSS	C>T	S>L	Unknown, N-terminal	1.7	1, 0	Neutral (2)
<i>MYH8</i>	Myosin 8 (myosin heavy chain 8)	17:10248420	WGSS	C>G	M>I	Actin-interacting protein domain	0	NA	Neutral (2)
<i>PALB2</i>	Partner and localizer of BRCA2	16:23559936	WGSS	T>G	E>A	N-terminal prefolding	13.0	NA	Amplification (4)
<i>PKDREJ</i>	Polycystic kidney disease and receptor for egg-jelly-related protein	22:45035285	WGSS	C>G	E>Q	Unknown	0.1	NA	Gain (3)
<i>RASEF</i>	RAS and EF-hand domain-containing protein	9:84867250	WTSS	G>A	S>L	EF-hand Ca ²⁺ -binding motif	65.0	3, 2	Gain (3)
<i>RNASEH2A</i>	Ribonuclease H2 subunit A (EC 3.1.26.4)	19:12785252	WGSS, WTSS	G>A	R>H	Unknown, C-terminal	5.3	2, 2	Neutral (2)
<i>RNF220</i>	RING finger protein C1orf164	1:44650831	WGSS	G>A	D>N	Unknown, N-terminal	16.1	NA	Neutral (2)
<i>SP1</i>	Transcription factor Sp1	12:52063157	WGSS	G>C	E>Q	Glu-rich N-terminal domain	57.3	40, 10 (Q < 0.01)	Amplification (4)
<i>USP28</i>	Ubiquitin carboxyl-terminal hydrolase 28	11:113185109	WGSS, WTSS	C>T	D>N	Unknown	12.5	3, 7	Gain (3)
<i>C11orf10</i>	UPF0197 transmembrane protein C11orf10	11:61313958	WGSS	G>A	T>I	Transmembrane domain	28.9	13, 3	Amplification (4)
<i>THRSP</i>	Thyroid hormone-inducible hepatic protein	11:77452594	WGSS	C>T	R>C	Unknown	0.3	NA	Gain (3)
<i>SCEL</i>	Sciellin	13:77076497	WGSS	A>G	K>R	Unknown	0.3	1, 0	Gain (3)
<i>SLC24A4</i>	Na ⁺ /K ⁺ /Ca ²⁺ -exchange protein 4	14:92018836	WGSS	G>A	V>I	Transmembrane domain	1.2	1, 0	Amplification (4)
<i>COL1A1</i>	Collagen alpha-1(I) chain precursor	17:45625043	WGSS	C>T	G>D	Pro-rich domain	80.0	24, 0 (Q < 0.01)	Amplification (4)
<i>KIAA1772</i>	GREB1-like protein	18:17278222	WGSS	A>G	D>G	Unknown	2.8	4, 1	Neutral (2)
<i>CCDC117</i>	Coiled-coil domain-containing protein 117	22:27506951	WGSS	G>C	K>N	Unknown	12.9	2, 0	Neutral (2)
<i>RP1-3210.10</i>	Novel protein	22:43140252	WGSS	G>C	E>Q	Unknown	0	NA	Gain (3)
<i>MORC1</i>	MORC family CW-type zinc finger protein 1	3:110271286	WGSS	G>A	A>V	Coiled-coil	0.1	NA	Gain (3)
<i>SNX4</i>	Sorting nexin 4	3:126721688	WGSS	C>T	D>N	Unknown, N-terminal	43.4	NA	Gain (3)
<i>LEPREL1</i>	Prolyl 3-hydroxylase 2 precursor (EC 1.14.11.7)	3:191172415	WGSS	T>C	E>G	Hydroxylase domain	1.1	NA	Gain (3)
<i>WDR59*</i>	WD repeat-containing protein 59	16:73500342	WTSS	C>T	M>I	Unknown, C-terminal	17.3	6, 5	Neutral (2)

Omnibus table showing the features associated with the 32 Sanger amplicon-validated non-synonymous somatic mutations from the WGSS-PE and WTSS-PE libraries. Mutation positions are on the basis of reference genome hg18. The nucleotide substitutions are shown as reference>variant. The amino acid change is shown as reference>variant amino acid. If the mutation occurs in a recognized protein domain or motif this is shown. The transcript expression level in WTSS-PE reads is shown as the mean number of reads supporting each position in the transcript. The allelic expression bias column shows the number of reference (R), non-reference (NR) reads in the WTSS-PE library at the mutated position. Three transcripts (*CHD3*, *SP1* and *COL1A1*) show significant expression bias (annotated with Q < 0.01, Supplementary Methods) in favour of the reference allele; however, none of the heterozygous somatic mutations were biased in favour of the non-reference allele. The expression of *HAUS3* is predominantly non-reference as expected for a homozygous allele. The HMM state classifier of copy number for the genomic region encompassing each mutation position is shown in the last column, as state (state number). C-terminal, carboxy-terminal; N-terminal, amino-terminal.

*Genes showing alternative splicing.

germline alleles and 32 were revealed as non-synonymous coding somatic point mutations (Table 2). Of the 32 somatic mutations, 30 were present in WGSS-PE and/or WTSS-PE, whereas two were detected from the WTSS library sequence alone (Table 2). None of the 32 genes were found in common with the CAN breast genes⁶, which were discovered from ER⁻ cell lines. Eleven genes appear in the current release of COSMIC⁷ (*CHD3*, *SP1*, *PALB2*, *ERBB2*, *USP28*, *KLHL4*, *CDC6*, *KIAA1468*, *RNF220*, *COL1A1* and *SNX4*) but with mutations at different positions. We examined the population frequency of the somatic mutation positions for *PALB2*, *ERBB2*, *USP28*, *CDC6*, *CHD3*, *HAUS3* (previously known as *C4orf15*), *SP1*, *KIAA1468* and *DLG4* in a further 192 breast cancers (Supplementary Methods; 112 lobular, 80 ductal). None of these 192 breast cancers showed identical mutations to those described here; however, 3 out of 192 cases (2 lobular, 1 ductal) contained neighbouring non-synonymous variants/deletions affecting the *ERBB2* kinase domain (Supplementary Fig. 4). Interestingly, 2 out of 192 cases (both lobular) contained two different heterozygous truncating variants in *HAUS3*: chr4:2203685 G>T on minus strand, GAG>TAG (Glu>stop), and chr4:2203483 C>G on minus strand, TCA>TGA (Ser>stop) (Supplementary Fig. 5). Notably, *HAUS3* is a member of the recently described^{8–10} multiprotein augmin complex, the function of which is required for genome stability mediated by appropriate kinetochore attachment and centrosome morphogenesis.

To determine how many of the somatic non-synonymous coding sequence mutations were already present at diagnosis 9 years earlier, we next examined genomic DNA from the primary tumour directly, by a single molecule frequency counting experiment (Supplementary Methods)⁴. Twenty-eight of the 32 mutations yielded amplicons compatible with Illumina sequencing (Supplementary Methods), and two extra mutations were sampled by Sanger sequencing

(Supplementary Fig. 5). As controls we selected 36 heterozygous germline SNVs at random. The PCR amplicons for known germline and somatic mutations were sequenced on an Illumina device. After alignment, the observed counts of reference and non-reference bases at the target position were compared using the Binomial exact test. To calibrate the expected mean of the Binomial distribution, we used the non-reference allele frequency from positions -5 to +5 surrounding (but not including) the target position (Supplementary Table 6a, b), where only reference bases should be called. Unequal segmental amplification/deletion in the genome may contribute to a departure from the theoretical ratio of 0.5 for a heterozygous allele. As a result, amplicons from heterozygous germline alleles showed occasional measured frequencies of between 0.2 and 0.8 in both the primary and metastatic tumour DNA (Table 3 and Supplementary Table 7), but with a modal frequency around 0.5, as expected. In the metastatic genomic DNA the somatic mutations showed frequencies of between 0.2 and 0.79 (Table 3). Notably, the somatic coding mutation positions examined in the primary tumour showed three patterns of abundance: prevalent, rare and undetectable (Table 3). Mutations in *ABCB11*, *PALB2* and *SLC24A4* were detected at prevalent frequencies for heterozygous mutations (≥ 0.2 , the lowest value seen for known germline alleles) given a 73% tumour content. The frequency of the mutation in *HAUS3* was 0.79, consistent with it being a prevalent homozygous mutation, also confirmed by Sanger sequencing (Supplementary Fig. 5). Sanger amplicon sequencing showed that the *SNX4* somatic mutation was also present in the primary tumour, whereas the *KIAA1772* (also known as *GREB1L*) mutation was not. Six mutations (*KIF1C*, *USP28*, *MORC1*, *MYH8*, *KIAA1468* and *RNASEH2A*) showed statistically significant ($P < 0.01$, Binomial exact test) intermediate frequencies of between 1% and 13% (Table 3), suggesting that these mutations were

Table 3 | Frequency of germline and somatic alleles in the metastatic and primary genomes

Position	Locus	R	NR	Primary depth	Primary NR ratio	Primary P value	Primary status	Metastasis depth	Metastasis NR ratio	M	Copy number classification (HMM state)
4:2203607	<i>HAUS3</i>	C	T	5700	0.5472	0.0000	Dominant	762	0.7874	S	Neutral (2)
16:23559936	<i>PALB2</i>	T	G	115	0.4957	0.0000	Dominant	669	0.4350	S	Amplification (4)
2:169497197	<i>ABCB11</i>	C	T	506	0.3261	0.0000	Dominant	959	0.3691	S	Amplification (4)
14:92018836	<i>SLC24A4</i>	G	A	13347	0.2341	0.0000	Dominant	13670	0.3518	S	Amplification (4)
17:10248420	<i>MYH8</i>	C	G	10657	0.1353	0.0000	Subdominant	1797	0.5932	S	Neutral (2)
3:110271286	<i>MORC1</i>	G	A	24572	0.0468	0.0000	Subdominant	32273	0.4107	S	Gain (3)
17:4848025	<i>KIF1C</i>	G	A	8587	0.0107	0.0000	Subdominant	2272	0.3077	S	Neutral (2)
11:113185109	<i>USP28</i>	C	T	6654	0.0095	0.0000	Subdominant	1387	0.4484	S	Gain (3)
18:58076768	<i>KIAA1468</i>	G	A	719	0.0083	0.0020	Subdominant	1056	0.3059	S	Neutral (2)
19:12785252	<i>RNASEH2A</i>	G	A	6537	0.0029	0.0276	Subdominant	1497	0.2806	S	Neutral (2)
4:106982671	<i>GSTCD</i>	G	T	7273	0.0008	0.9885	Absent	2208	0.2174	S	Neutral (2)
17:35701114	<i>CDC6</i>	G	T	4894	0.0008	0.9733	Absent	4208	0.3577	S	Amplification (4)
17:7751231	<i>CHD3</i>	G	A	9665	0.0007	0.9981	Absent	1737	0.2671	S	Neutral (2)
4:155726802	<i>FGA</i>	C	T	5756	0.0007	0.9911	Absent	2287	0.2755	S	Gain (3)
17:7052251	<i>DLG4</i>	G	A	4383	0.0007	0.9835	Absent	706	0.3272	S	Neutral (2)
3:37267947	<i>GOLGA4</i>	G	T	13051	0.0006	0.9999	Absent	3262	0.2235	S	Gain (3)
9:84867250	<i>RASEF</i>	G	T	1690	0.0006	0.9500	Absent	796	0.3656	S	Gain (3)
17:35133783	<i>ERBB2</i>	C	A	3736	0.0005	0.9899	Absent	1722	0.3612	S	Amplification (4)
X:86659878	<i>KLHL4</i>	C	T	6561	0.0005	0.9993	Absent	977	0.3153	S	Neutral (2)
3:191172415	<i>LPREL1</i>	T	C	11963	0.0004	1.0000	Absent	8381	0.2148	S	Gain (3)
16:73500342	<i>WDR59</i>	C	T	4846	0.0004	0.9982	Absent	1396	0.2629	S	Neutral (2)
1:44650831	<i>RNF220</i>	G	A	8160	0.0004	0.9999	Absent	967	0.2203	S	Neutral (2)
22:45035285	<i>PKDREJ</i>	C	T	6674	0.0003	0.9999	Absent	1230	0.3366	S	Gain (3)
11:61313958	<i>C11ORF10</i>	G	A	116705	0.0003	1.0000	Absent	14354	0.4651	S	Amplification (4)
12:52063157	<i>SP1</i>	G	T	7732	0.0003	1.0000	Absent	2011	0.2193	S	Amplification (4)
11:77452594	<i>THRSP</i>	C	T	24219	0.0002	1.0000	Absent	40652	0.4750	S	Gain (3)
17:45625043	<i>COL1A1</i>	C	A	26343	0.0001	1.0000	Absent	32259	0.2543	S	Amplification (4)
13:77076497	<i>SCEL</i>	A	G	49	0.0000	1.0000	Absent	187	0.5722	S	Gain (3)
19:9314428	—	A	G	176	0.5057	0.0000	Present	321	0.4953	G	Neutral (2)
4:130144460	—	A	T	2020	0.2188	0.0000	Present	2081	0.3099	G	Neutral (2)
8:27835012	—	G	A	13587	0.8602	0.0000	Present	10781	0.6667	G	Deletion (1)
6:32908543	—	C	T	4718	0.7484	0.0000	Present	16370	0.4897	G	Amplification (4)
20:43363061	—	G	A	5950	0.5249	0.0000	Present	5540	0.5049	G	Amplification (4)
4:8672089	—	G	A	381	1.0000	0.0000	Present	2850	0.8032	G	Gain (3)
16:1331138	—	C	T	677	0.4963	0.0000	Present	554	0.6245	G	High-level amplicon (5)

Only 7 germline alleles are shown, the full list is in Supplementary Table 7. The genome positions are shown as chr:coordinate. The primary read depth represents the number of reads. Binomial exact P values were calculated using a Binomial exact test. R, reference base; NR, non-reference base. Primary status indicates whether the variant was present, subdominant or absent in the primary tumour. Column M denotes somatic (S) or germline (G) single nucleotide variants in the metastasis. HMM state refers to the metastasis.

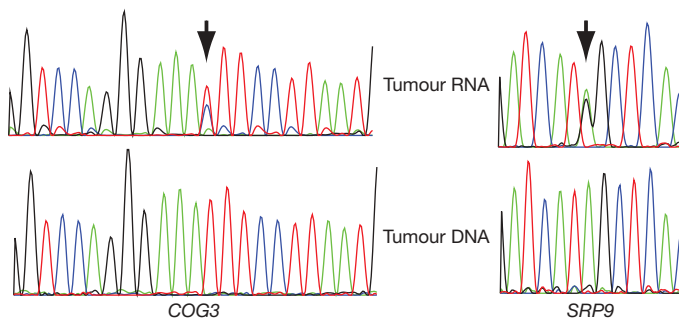


Figure 1 | RNA editing in COG3 and SRP9. Sanger sequence traces from the non-synonymous editing positions in COG3 and SRP9. The editing position is arrowed. Top trace is tumour RNA, bottom trace tumour DNA. The editing positions were confirmed with reverse strand reads (not shown).

restricted to minor subclones of tumour cells. The remaining 19 out of 30 of the somatic coding mutations were not detected in the primary tumour DNA. Thus, significant heterogeneity in tumour somatic mutation content existed in the primary tumour at diagnosis. In contrast with the recently reported sequence of cytogenetically normal acute myeloid leukaemia (AML) tumour⁴, significant evolution of coding mutational content occurred between primary and metastasis. It is unknown whether the 19 mutations present in the metastasis, but not detected in the primary, were a consequence of radiation therapy or innate tumour progression.

We also examined how the transfer of information from the nuclear genome to proteins was modified by alternative splicing (Supplementary Table 8 and Supplementary Fig. 6), biased allelic expression (Supplementary Table 9) and RNA editing. At the single nucleotide level, RNA-editing enzymes (which can be regulated by oestrogens¹¹) may also recode transcripts resulting in a proteome divergent from the genome^{12–15}. Interestingly, the ADAR enzyme—one of the principal RNA-editing enzymes that mediates A→I(G) edits—was one of the top 5% of genes expressed (145.6 reads per base, Supplementary Table 10), and the only editing enzyme expressed at a high level. We searched for potential editing events (Methods) and found 3,122 candidate edits in 1,637 gene loci (Supplementary Table 11). Some 526 out of 3,122 candidate edits are non-synonymous changes and 232 are synonymous changes (with the remainder affecting untranslated regions). We revalidated independently (Supplementary Methods) by Sanger sequencing 75 editing events in 12 gene loci from the lobular metastasis (Supplementary Table 12 and see trace data at <http://molonc.bccrc.ca/>). Two genes, COG3 and SRP9 (Fig. 1), showed confirmed high frequency non-synonymous transcript editing, resulting in variant protein sequences. These observations emphasize the importance of integrating RNA-seq data with tumour genomes in assessing protein variation.

The coding mutation landscape of breast cancers has, so far, been mostly determined from ER⁺ metastatic cell lines/samples^{6,16}, and has suggested the presence of large numbers of passenger events as well as drivers. Our results show the importance of sequencing samples of tumour cell populations early as well as late in the evolution of tumours, and of estimating allele frequency in tumour genomes. Our observations suggest that the sequencing of primary breast cancers and pre-invasive malignancy may reveal significantly fewer candidates for tumour initiating mutations.

METHODS SUMMARY

Paired-end reads were assigned quality scores and aligned to the reference genome (hg18) using Maq¹⁷ (Supplementary Methods and Supplementary Fig. 2). For identification of SNVs we used a simple Binomial mixture model, SNVMix (Supplementary Appendix 1), which assigns a probability to each base position as homozygous reference (aa), heterozygous non-reference (ab) and homozygous non-reference (bb), based on the occurrence of reference (hg18) and

non-reference bases at each aligned position. This model was calibrated initially, using high confidence allele calls from Affymetrix SNP6.0 hybridization of tumour and normal DNA. We estimated the receiver operating characteristic (ROC) performance (Supplementary Fig. 8) and determined that an SNVMix threshold of $P = 0.77$ for (ab) or (bb) for a non-reference call would yield a false discovery rate (FDR) of 1%. For the RNA-seq library, a threshold of $P = 0.53$ was used (Supplementary Fig. 8; FDR = 0.01) to call non-reference positions. Non-reference positions were then filtered for known variants against the sources of germline variation, the single nucleotide polymorphism database (dbSNP) and the completed individual genomes^{18,19} (Supplementary Table 2). Saturation of the libraries for SNV discovery was determined by random re-sampling (Supplementary Fig. 9 and Supplementary Methods). Segmental copy number was inferred with a hidden Markov model (HMM) method (Supplementary Table 4a, b and Supplementary Methods).

We searched for RNA-editing events by examining all very high confidence ($P(ab) + P(bb) > 0.9$) SNVMix predictions from the RNA-seq library of the metastatic tumour, that were not found with extreme confidence ($P(aa) > 0.99$, derived from the SNVMix receiver operating curve at FDR = 0.01) at the same positions in the metastatic tumour genome library.

Received 4 September; accepted 10 September 2009.

- Campbell, P. J. *et al.* Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nature Genet.* **40**, 722–729 (2008).
- Bentley, D. R. *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**, 53–59 (2008).
- Morin, R. *et al.* Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing. *Biotechniques* **45**, 81–94 (2008).
- Ley, T. J. *et al.* DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* **456**, 66–72 (2008).
- Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* **5**, 621–628 (2008).
- Wood, L. D. *et al.* The genomic landscapes of human breast and colorectal cancers. *Science* **318**, 1108–1113 (2007).
- Forbes, S. A. *et al.* The Catalogue of Somatic Mutations in Cancer (COSMIC). *Curr. Protoc. Hum. Genet.* Unit 10.11, doi:10.1002/0471142905.hg1011s57 (2008).
- Goshima, G., Mayer, M., Zhang, N., Stuurman, N. & Vale, R. D. Augmin: a protein complex required for centrosome-independent microtubule generation within the spindle. *J. Cell Biol.* **181**, 421–429 (2008).
- Meireles, A. M., Fisher, K. H., Colombini, N., Wakefield, J. G. & Ohkura, H. Wac: a new Augmin subunit required for chromosome alignment but not for centrosomal microtubule assembly in female meiosis. *J. Cell Biol.* **184**, 777–784 (2009).
- Lawo, S. *et al.* HAU5, the 8-subunit human Augmin complex, regulates centrosome and spindle integrity. *Curr. Biol.* **19**, 816–826 (2009).
- Pauklin, S., Sernandez, I. V., Bachmann, G., Ramiro, A. R. & Petersen-Mahrt, S. K. Estrogen directly activates AID transcription and function. *J. Exp. Med.* **206**, 99–111 (2009).
- Blow, M., Futreal, P. A., Wooster, R. & Stratton, M. R. A survey of RNA editing in human brain. *Genome Res.* **14**, 2379–2387 (2004).
- Athanasiadis, A., Rich, A. & Maas, S. Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome. *PLoS Biol.* **2**, e391 (2004).
- Maas, S., Kawahara, Y., Tamburro, K. M. & Nishikura, K. A-to-I RNA editing and human disease. *RNA Biol.* **3**, 1–9 (2006).
- Li, J. B. *et al.* Genome-wide identification of human RNA editing sites by parallel DNA capturing and sequencing. *Science* **324**, 1210–1213 (2009).
- Greenman, C. *et al.* Patterns of somatic mutation in human cancer genomes. *Nature* **446**, 153–158 (2007).
- Li, H., Ruan, J. & Durbin, R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res.* **18**, 1851–1858 (2008).
- Wang, J. *et al.* The diploid genome sequence of an Asian individual. *Nature* **456**, 60–65 (2008).
- Wheeler, D. A. *et al.* The complete genome of an individual by massively parallel DNA sequencing. *Nature* **452**, 872–876 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Eaves and M. Pollak for comments on earlier versions of the manuscript. We thank and acknowledge the patients of the BC Cancer Agency for donations of tumour tissues to the TTR-BREAST tumour banking program. S.A. is supported by a Canada Research Chair in Molecular Oncology, S.P.S., J.K., L.P., A.B. and T.P. are supported by Michael Smith Foundation for Health Research awards. R.D.M. is a Vanier scholar (CIHR). A.B. is also supported by an NSERC award, and L.P. by a CIHR award. We are grateful for platform support from CIHR, Genome Canada, Genome BC, Canada Foundation for Innovation and the Michael Smith Foundation for Health Research. The work was funded by the BC Cancer Foundation and the CBCF BC/Yukon chapter.

Author Contributions S.P.S. and R.D.M.: led the data analysis and wrote the manuscript. M.H.: oversaw the sequencing efforts. J.K., L.P., T.P., J.S., C.S., A.B.,

R.M. and T.S.: validation of variants. A.D.: primer design. K.G. and P.W.: establishment of TTR-BREAST tumour bank. K.T., R.G., R.A.H., S.J., M.S., G.L., A.E.T., R.V., G.A.T. and R.L.W.: bioinformatic analysis. G.T., D.H. and P.W.: sample selection and histological grading. Y.Z.: Illumina sequencing library preparation. C.C. and D.H.: data analysis and interpretation. S.A. and M.A.M.: conceived and oversaw the study and wrote the manuscript.

Author Information Genome sequence data have been deposited at the European Genotype Phenotype Archive (<http://www.ebi.ac.uk/ega>), which is hosted by the EBI, under accession number EGAS00000000054. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.A.M. (mmarra@bcgsc.ca) or S.A. (saparicio@bccrc.ca).

LETTERS

Direct RNA sequencing

Fatih Ozsolak¹, Adam R. Platt¹, Dan R. Jones¹, Jeffrey G. Reifenger¹, Lauryn E. Sass¹, Peter McInerney¹, John F. Thompson¹, Jayson Bowers¹, Mirna Jarosz¹ & Patrice M. Milos¹

Our understanding of human biology and disease is ultimately dependent on a complete understanding of the genome and its functions. The recent application of microarray and sequencing technologies to transcriptomics has changed the simplistic view of transcriptomes to a more complicated view of genome-wide transcription where a large fraction of transcripts emanates from unannotated parts of genomes^{1–7}, and underlined our limited knowledge of the dynamic state of transcription. Most of this broad body of knowledge was obtained indirectly because current transcriptome analysis methods typically require RNA to be converted to complementary DNA (cDNA) before measurements, even though the cDNA synthesis step introduces multiple biases and artefacts that interfere with both the proper characterization and quantification of transcripts^{8–18}. Furthermore, cDNA synthesis is not particularly suitable for the analysis of short, degraded and/or small quantity RNA samples. Here we report direct single molecule RNA sequencing without prior conversion of RNA to cDNA. We applied this technology to sequence femtomole quantities of poly(A)⁺ *Saccharomyces cerevisiae* RNA using a surface coated with poly(dT) oligonucleotides to capture the RNAs at their natural poly(A) tails and initiate sequencing by synthesis. We observed transcript 3' end heterogeneity and polyadenylated small nucleolar

RNAs. This study provides a path to high-throughput and low-cost direct RNA sequencing and achieving the ultimate goal of a comprehensive and bias-free understanding of transcriptomes.

cDNA-based transcriptome analysis approaches being used today exhibit several shortcomings that prevent us from understanding the real nature of transcriptomes and ultimately genome biology. Some of these limitations are: (1) the tendency of various reverse transcriptases (RT) to generate spurious second-strand cDNA due to their DNA-dependent DNA polymerase activities^{9,10,18}; (2) the generation of artefactual cDNAs due to template switching^{8,13,16,17} or contaminating DNA and primer-independent cDNA synthesis^{11,12}; and (3) the error-prone^{15,19} and inefficient nature of RTs yielding low quantities of cDNA. Furthermore, most RNA analysis technologies require the synthesis of not just the first strand cDNA but also a second strand cDNA that are both subjected to further ligation/amplification steps, introducing yet more biases. These limitations pose problems for the determination of RNA strandedness^{14,20}, the identification of chimaeric transcripts, quantification of RNA species, and the analysis of low quantity (<1 nanogram) or short RNA species, such as those obtained from formalin-fixed, paraffin-embedded tissue samples. Because almost all transcript analysis technologies in use today suffer from the limitations briefly summarized above, there is an ever-growing

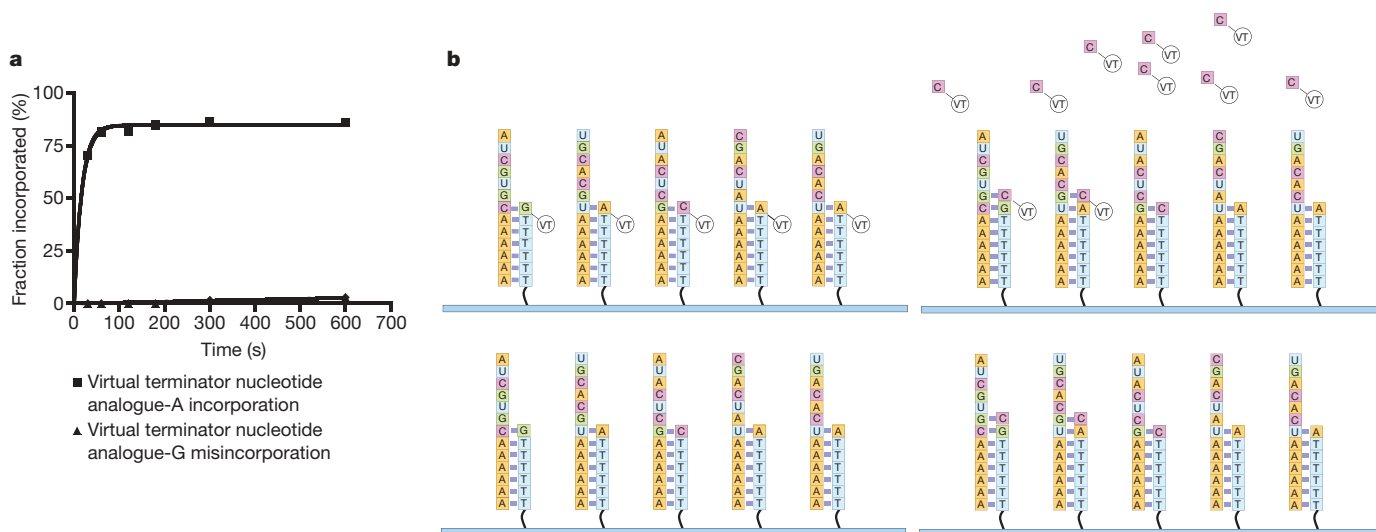


Figure 1 | DRS chemistry and sequencing steps. **a**, Under optimized conditions, polymerase exhibits fast correct nucleotide incorporation (VT-A) and slow misincorporation (VT-G) kinetics. **b**, DRS procedure. Top left: polyadenylated and 3'-blocked RNA is captured on surfaces coated with dT(50) oligonucleotide. A 'fill' step is performed with natural dTTP, and a 'lock' step with fluorescently labelled VT-A, -C and -G nucleotides. These steps correct for any misalignments that may be present in poly(A/T) duplexes, and ensure that the sequencing starts in the template rather than

the poly(A) tail. Imaging is performed to locate the template positions. Bottom left: chemical cleavage of the dye-nucleotide linker is performed to prepare the templates for nucleotide incorporation. Top right: incubation with one VT nucleotide and polymerase is performed, followed by imaging to locate the templates that incorporated the nucleotide. Bottom right: chemical cleavage of the dye allows the surface and RNA templates to be ready for the next nucleotide addition cycle.

¹Helicos BioSciences Corporation, One Kendall Square, Cambridge, Massachusetts 02139, USA.

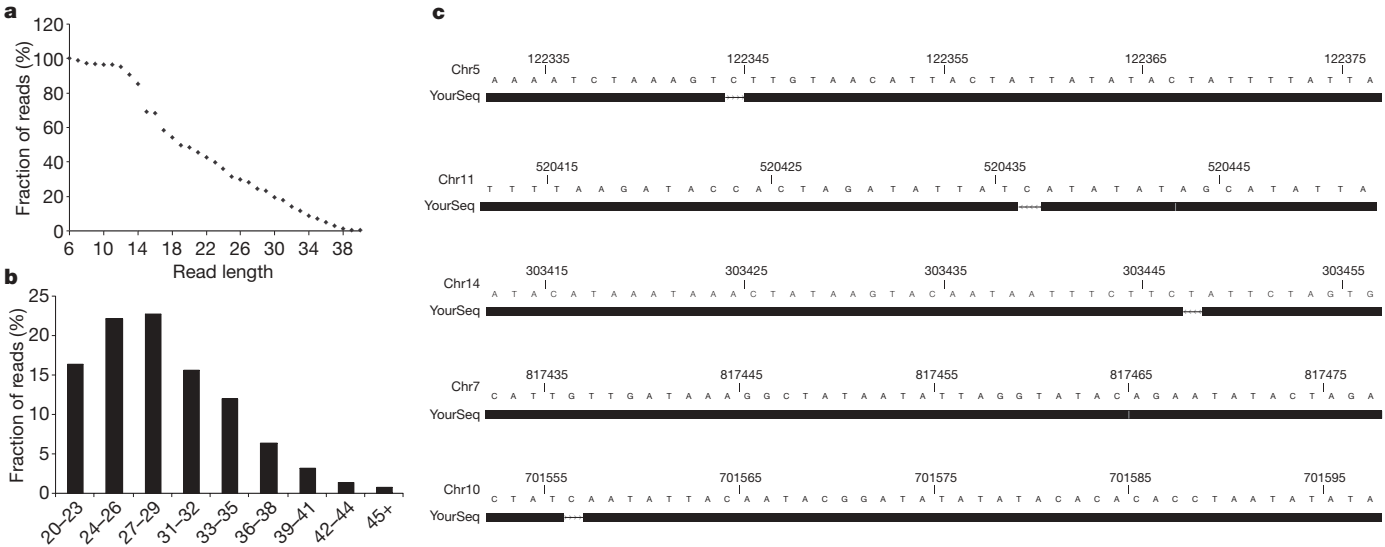


Figure 2 | DRS sequencing read-length statistics. **a**, Cumulative length distribution of reads obtained from oligoribonucleotides. *y* axis shows the fraction of reads at and above particular *x*-read lengths. **b**, Distribution of

read lengths greater than 20 nucleotides aligned to *S. cerevisiae* genome. **c**, Several DRS reads are aligned with BLAT and visualized using the UCSC genome browser.

need for a method that would not be subject to the difficulties associated with RT behaviour, amplification, ligation and other cDNA synthesis/sample manipulation steps. A method allowing a comprehensive and bias-free view of transcriptomes using minute quantities

of total RNA obtained from as few as one cell with no pre-treatment would stimulate great advances in the delineation of complex biological processes and be applicable across all biomedical research areas.

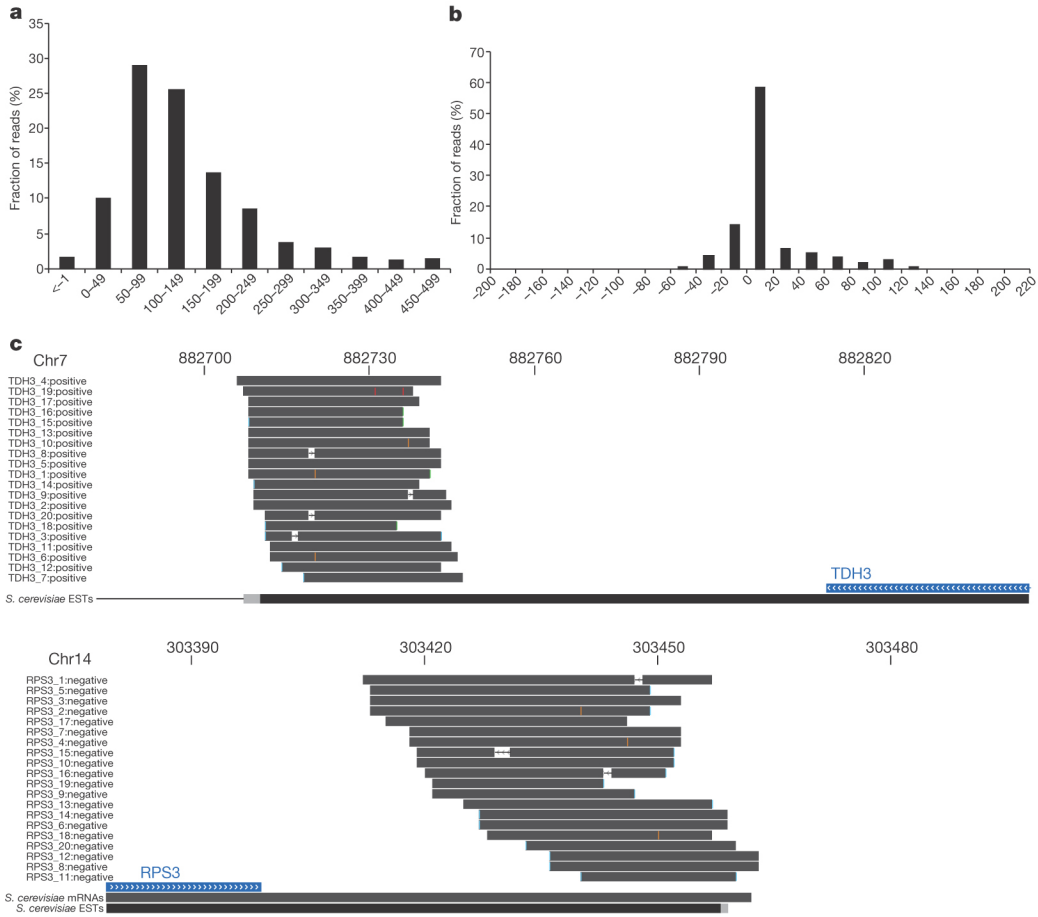


Figure 3 | DRS read distribution. **a**, **b**, Distance of aligned reads to *S. cerevisiae* coding sequence (**a**) and EST 3' ends (**b**). *y* axis shows the fraction of reads at particular distance intervals indicated on the *x* axis. Most reads (91%) are 300 nucleotides immediately downstream of annotated gene 3' ends. Reads aligning within the coding regions of transcripts are shown

with negative distance in **a**. **b** represents the distances between the DRS reads and the closest EST clones. **c**, DRS reads aligning to *TDH3* and *RPS3* are exemplified. The alignment direction of each read (black bars) is indicated as 'positive' or 'negative'.

Here we report the successful development of direct RNA sequencing (DRS), allowing massively parallel sequencing of RNA molecules directly without prior synthesis of cDNA or the need for ligation/amplification steps. DRS represents an extension of single-molecule DNA sequencing technology (tSMS)^{21,22} that relies on the stepwise synthesis and direct imaging of billions of single DNA strands on a planar surface. The sequencing-by-synthesis reaction is performed using a modified polymerase and proprietary fluorescent nucleotide analogues, called Virtual Terminator nucleotides (VT), that contain a fluorescent dye and chemically cleavable groups that allow step-wise sequencing. The first step for the development of DRS was the identification of an optimal polymerase, VT nucleotide analogues and buffer combination. Several DNA-dependent DNA polymerases have previously been shown to have reverse transcriptase activity^{23–25}; we therefore tested DNA polymerases in addition to known RTs. After screening studies performed in solution, we identified conditions with satisfactory reaction kinetics (Fig. 1a) that could be attempted in a single-molecule sequencing system. The DRS procedure is summarized in Fig. 1b. Briefly, *Escherichia coli* poly(A) polymerase I (PAPI) is used to generate an A tail on 3' ends of RNA molecules. The control of the A-tail length and the 3' end blocking is performed by introducing 3'-deoxyATP to the polyadenylation reaction shortly after the start of

the tailing reaction, generating an A-tail of ~150 nucleotides. The blocking step is performed to prevent 'downward' nucleotide additions to the 3' end of the template during the sequencing process. For RNA species containing poly(A) tails, such as mRNAs, poly(A) tailing is not required; only 3' blocking is needed. Polyadenylated and 3'-blocked RNAs are hybridized to poly(dT)-coated surfaces. To begin sequencing at the unique region adjacent to the poly(A) tail, each RNA molecule is 'filled' in with dTTP and polymerase, and then 'locked' in position with VT-A, -C and -G addition, stopping subsequent nucleotide additions. After washing away the unincorporated dye-labelled nucleotides, images are taken, and then the fluorescent dye and inhibitor are cleaved off the incorporated nucleotide, rendering it suitable for additional rounds of incorporation. Each molecule is then provided the opportunity to extend (alternating C, T, A or G) followed by rinsing, imaging and cleavage. Repeating this cycle many times provides a set of images that are aligned and then used to generate sequence information for each individual RNA molecule with real-time image processing.

We first used chemically synthesized 40-mer RNA oligoribonucleotides as a model system to develop and optimize DRS chemistry. After sequencing on a prototype sequencer with 120 cycles of alternating VT-C, -T, -A or -G additions, we aligned the resultant sequence reads to the input oligonucleotide reference sequences and observed 48.5%

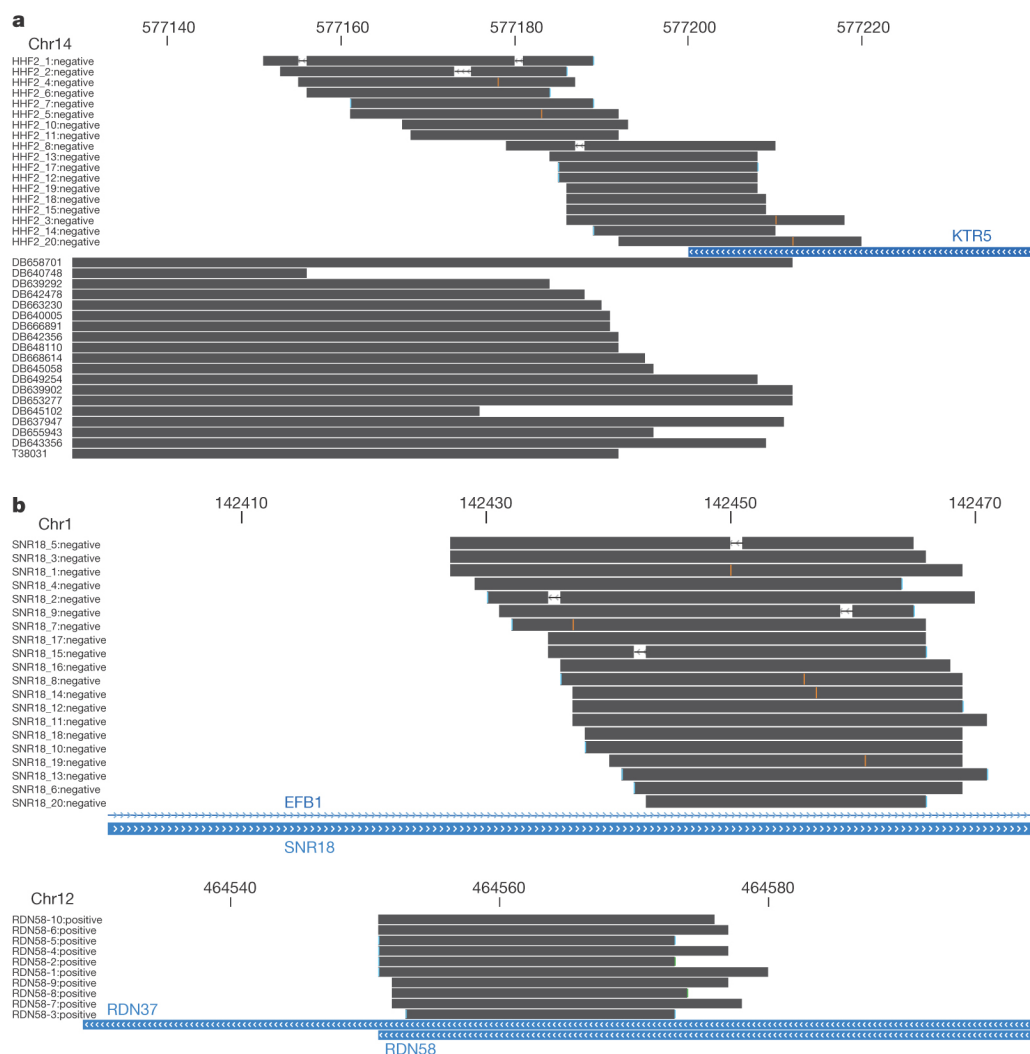


Figure 4 | *S. cerevisiae* poly(A)⁺ RNA DRS suggests overlapping transcription units and polyadenylated snoRNA and rRNA species. a, DRS reads in this region are aligned in the reverse direction, suggesting their origination from the *HHF2* transcript (not shown) located ~200 nucleotide upstream relative to annotated *KTR5* 3' end and transcribed in the forward direction. Three reads extend into the *KTR5* coding sequence, suggesting

that the *HHF2* transcription can extend into *KTR5* coding sequence. **b**, Figure exemplifies randomly selected reads aligning to intronic *SNR18* (top panel) and 5.8S rRNA (bottom panel). Note that the reads are aligning within the 3' ~40-nucleotide annotated regions of snoRNAs and rRNAs, suggesting their polyadenylation during or after their maturation.

of aligned reads to have a sequence length of at least 20 nucleotides, with the longest perfect match (no errors) being 38 nucleotides (Fig. 2a and Supplementary Fig. 5). In terms of total strand yield, DRS was efficient, providing on average 972 reads per 1,000 μm^2 flow cell surface area compared to $\sim 1,100$ for tSMS DNA sequencing in similar conditions. Because the sequencing process relies on incubating the templates with one base at a time, all errors are single base errors in the form of deletions (failure to detect incorporation), insertions (for example, failure to rinse VT analogues from the flow cell between each addition cycle) and substitutions. Total raw base error rate for DRS is currently approximately 4%, dominated by missing base errors (2–3%), whereas the insertion rate is 1–2% and the substitution error rate is 0.1–0.3%. Although further improvements in error rates are in progress, the read lengths and error rates achieved here are sufficient to allow the use of standard computational methods to align sequences to reference transcriptomes and genomes.

We then sequenced *Saccharomyces cerevisiae* poly(A)⁺ RNA with DRS. Because this RNA sample already contains a natural, pre-existing poly(A) tail, no additional tailing was needed. Two femtomoles (~ 2 nanograms) of 3' end-blocked yeast poly(A)⁺ RNA was hybridized to dT(50) flow cells with no additional sample preparation procedures. One hundred and twenty sequencing cycles were performed on a prototype sequencing system over 3 days. RNA stability remained at high levels during the run, as demonstrated by the relatively constant number of nucleotides added per addition cycle to RNA templates (Supplementary Fig. 2). The sequence run generated 41,261 reads greater than 20 nucleotides, of which 19,501 reads (48.4%) aligned to the yeast genome using the BLAT algorithm²⁶. The average aligned read length was 28.7 nucleotides, with the longest perfect match aligned read being 50 nucleotides (Fig. 2b, c). Of the aligned reads 91% were within 400 nucleotides downstream of annotated yeast gene 3' open reading frame sequence ends (Fig. 3a). Such a wide distribution is expected, as yeast 3' gene annotations mark mostly the coding sequence end point rather than the polyadenylation site. As expected, the alignment orientation of these sequences was in the opposite direction to the known gene transcription direction. This is a result of using unmodified, intact yeast poly(A)⁺ RNA without any additional sample preparation steps, and therefore, the sequence read matches the direction opposite to that of the transcript. Because the 3' ends of yeast protein-coding genes are not well annotated, we compared our findings to the yeast expressed sequence tags (EST) database as well. As exemplified in Fig. 3b and c, our data are supported by the EST data, with most of our reads being in close proximity to EST 3' ends, aligned in the direction opposite to transcription. The reads that did not align to proximal 3' ends of yeast genes or EST clones were caused by their localization beyond the 500 nucleotide downstream regions examined, at the 3' ends of potentially transcriptionally active retrotransposons or at the 3' ends of transcripts classified as dubious (Supplementary Fig. 3). Comparison of the DRS read localizations to the transcript 3' ends identified previously⁵ using high-throughput cDNA sequencing revealed a high concordance, with 81% of the DRS reads being within ± 20 nucleotides of their 3'-end annotations (Supplementary Fig. 6). We observed transcripts extending into the coding sequence of neighbouring genes (Fig. 4a), some of which were supported by the available EST data and as described⁵. Interestingly, $\sim 2\%$ of the total reads aligning within the coding regions of transcripts (Fig. 3a) were from the ribosomal RNAs (rRNAs) and a portion of small nucleolar RNAs (snoRNAs). Mature forms of these RNAs are produced from longer precursor RNAs through cleavage steps. Our observation that DRS reads map to the 3' end of the mature snoRNAs and rRNAs indicates that at least a fraction of snoRNAs^{27,28} and rRNAs²⁹ can be polyadenylated post-transcriptionally, possibly during their 3'-end processing and/or RNA quality control steps²⁸ (Fig. 4b and Supplementary Fig. 7). We independently validated the 3' polyadenylation site heterogeneity and the existence of polyadenylated snoRNAs by amplifying 3' polyadenylation sites with polymerase chain reaction (PCR) in a manner preserving the variability

in the 3' ends, followed by sequencing of the PCR products with tSMS DNA sequencing to identify the 3' polyadenylation sites (Supplementary Fig. 11). Our data add further support to the suggestion that many yeast genes possess a heterogeneous set of 3' ends for genes⁵.

The simplicity of the DRS sample preparation steps presented here, the requirement for only femtomole quantities of RNA and the potential of DRS to eliminate biases introduced by cDNA synthesis, end repair, ligation and amplification procedures will be useful for applications requiring minute quantities of RNA and/or short RNA species that are challenging for analysis with existing cDNA-based methodologies. This ability, combined with further improvements in DRS sample preparation, single molecule sequencing surface capture, throughput and computational tools, will ultimately allow us to understand and quantify the 'true' nature of transcriptomes in a high-throughput, low-cost and bias-free manner.

METHODS SUMMARY

RNA oligoribonucleotide templates were obtained from IDT. *S. cerevisiae* poly(A)⁺ RNA was obtained from Clontech. Polyadenylation of oligoribonucleotides was performed by using a poly(A) tailing kit (Ambion). 3'-deoxyATP (cordycepin triphosphate, Jena Biosciences) was introduced 10 min after the initiation of the polyadenylation reaction for 3'-end blocking and tail length limitation. DRS reads obtained are listed in Supplementary Table 4.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 18 May; accepted 5 August 2009.

Published online 23 September 2009.

- Denoeuf, F. *et al.* Annotating genomes with massive-scale RNA sequencing. *Genome Biol.* **9**, R175 (2008).
- Kapranov, P., Willingham, A. T. & Gingeras, T. R. Genome-wide transcription and the implications for genomic organization. *Nature Rev. Genet.* **8**, 413–423 (2007).
- Marioni, J. C., Mason, C. E., Mane, S. M., Stephens, M. & Gilad, Y. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res.* **18**, 1509–1517 (2008).
- Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* **5**, 621–628 (2008).
- Nagalakshmi, U. *et al.* The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* **320**, 1344–1349 (2008).
- Sultan, M. *et al.* A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science* **321**, 956–960 (2008).
- Wilhelm, B. T. *et al.* Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature* **453**, 1239–1243 (2008).
- Cocquet, J., Chong, A., Zhang, G. & Veitia, R. A. Reverse transcriptase template switching and false alternative transcripts. *Genomics* **88**, 127–131 (2006).
- Gubler, U. Second-strand cDNA synthesis: classical method. *Methods Enzymol.* **152**, 325–329 (1987).
- Gubler, U. Second-strand cDNA synthesis: mRNA fragments as primers. *Methods Enzymol.* **152**, 330–335 (1987).
- Haddad, F. *et al.* Regulation of antisense RNA expression during cardiac MHC gene switching in response to pressure overload. *Am. J. Physiol. Heart Circ. Physiol.* **290**, H2351–H2361 (2006).
- Haddad, F., Qin, A. X., Giger, J. M., Guo, H. & Baldwin, K. M. Potential pitfalls in the accuracy of analysis of natural sense-antisense RNA pairs by reverse transcription-PCR. *BMC Biotechnol.* **7**, 21 (2007).
- Mader, R. M. *et al.* Reverse transcriptase template switching during reverse transcriptase-polymerase chain reaction: artificial generation of deletions in ribonucleotide reductase mRNA. *J. Lab. Clin. Med.* **137**, 422–428 (2001).
- Perocchi, F., Xu, Z., Clauder-Munster, S. & Steinmetz, L. M. Antisense artifacts in transcriptome microarray experiments are resolved by actinomycin D. *Nucleic Acids Res.* **35**, e128 (2007).
- Roberts, J. D. *et al.* Fidelity of two retroviral reverse transcriptases during DNA-dependent DNA synthesis *in vitro*. *Mol. Cell. Biol.* **9**, 469–476 (1989).
- Roy, S. W. & Irimia, M. When good transcripts go bad: artifactual RT-PCR 'splicing' and genome analysis. *Bioessays* **30**, 601–605 (2008).
- Roy, S. W. & Irimia, M. Intron mis-splicing: no alternative? *Genome Biol.* **9**, 208 (2008).
- Spiegelman, S. *et al.* DNA-directed DNA polymerase activity in oncogenic RNA viruses. *Nature* **227**, 1029–1031 (1970).
- Varadaraj, K. & Skinner, D. M. Denaturants or cosolvents improve the specificity of PCR amplification of a G + C-rich DNA using genetically engineered DNA polymerases. *Gene* **140**, 1–5 (1994).
- Wu, J. Q. *et al.* Systematic analysis of transcribed loci in ENCODE regions using RACE sequencing reveals extensive transcription in the human genome. *Genome Biol.* **9**, R3 (2008).

21. Braslavsky, I., Hebert, B., Kartalov, E. & Quake, S. R. Sequence information can be obtained from single DNA molecules. *Proc. Natl Acad. Sci. USA* **100**, 3960–3964 (2003).
 22. Harris, T. D. *et al.* Single-molecule DNA sequencing of a viral genome. *Science* **320**, 106–109 (2008).
 23. Karkas, J. D., Stavrianopoulos, J. G. & Chargaff, E. Action of DNA polymerase I of *Escherichia coli* with DNA-RNA hybrids as templates. *Proc. Natl Acad. Sci. USA* **69**, 398–402 (1972).
 24. Rüttimann, C., Cotoras, M., Zaldívar, J. & Vicuna, R. DNA polymerases from the extremely thermophilic bacterium *Thermus thermophilus* HB-8. *Eur. J. Biochem.* **149**, 41–46 (1985).
 25. Stenesh, J., Roe, B. A. & Snyder, T. L. Studies of the deoxyribonucleic acid from mesophilic and thermophilic bacteria. *Biochim. Biophys. Acta* **161**, 442–454 (1968).
 26. Kent, W. J. BLAT—the BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002).
 27. Kim, M. *et al.* Distinct pathways for snoRNA and mRNA termination. *Mol. Cell* **24**, 723–734 (2006).
 28. Grzechnik, P. & Kufel, J. Polyadenylation linked to transcription termination directs the processing of snoRNA precursors in yeast. *Mol. Cell* **32**, 247–258 (2008).
 29. Slomovic, S., Laufer, D., Geiger, D. & Schuster, G. Polyadenylation of ribosomal RNA in human cells. *Nucleic Acids Res.* **34**, 2966–2975 (2006).
- Supplementary Information** is linked to the online version of the paper at www.nature.com/nature.
- Acknowledgements** We thank K. Kerouac, P. Kapranov, L. Kung, C. Hart and D. Lipson for technical assistance and discussions.
- Author Contributions** F.O. conceived the project, designed the experimental plan, coordinated the studies and analysed the data. F.O., J.B., L.E.S. and P.M. performed the enzyme kinetics assays. F.O., D.R.J., A.R.P. and J.G.R. did the sequencing experiments. J.F.T. and M.J. provided experimental reviews. F.O. and P.M.M. wrote the manuscript, which was reviewed by all authors.
- Author Information** Sequencing data sets described in this study have been deposited at the National Center for Biotechnology Information (NCBI) Short Read Archive (SRA), accession no SRA 009023. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Correspondence and requests for materials should be addressed to F.O. (fatihozsolak@gmail.com) or P.M.M. (pmilos@helicosbio.com).

METHODS

Polymerase kinetics assay. Incorporations of VT nucleotide analogues in an RNA-template-directed manner by various enzyme and buffer combinations were screened by designing four 50-mer oligoribonucleotides (5'-UUCUUUU GCCUCUUUCGNCAGGGCAGAGGAUGGAUGCAAGGAUAAGUGGA-3'); the 5' 25-nucleotide sequence of the oligoribonucleotides being complementary to a 25-mer 5'-rhodamine-labelled DNA oligo (5'-TCCACTTATCCTTGCAT CCATCCTCTGCCCTG-3'), and the 26th nucleotide (denoted as N above) on the oligoribonucleotides being each of the four nucleotides. After hybridizing the 25-mer DNA oligonucleotide to RNA templates at 65 °C for 5 min in nuclease-free water, followed by incubation on ice for 2 min, the selected enzyme/buffer/VT combinations were added to the RNA-DNA hybrid mix. The reaction was stopped at different time points by an EDTA-quench and kinetics were measured by observing the lengthening of the 5'-rhodamine-labelled DNA oligonucleotide using capillary electrophoresis (ABI 3730 DNA Analyzer, Applied Biosystems). This assay allowed us to observe the kinetics of VT nucleotide incorporation into the 3' end of the DNA primer in an RNA-template-directed manner. All oligonucleotides and oligoribonucleotides were ordered from IDT.

Sample preparation for DRS. RNA oligoribonucleotide templates were ordered from IDT. The sequences were Oligo 1, 5'-AGAGUCCCAUCCACCAUCAU CACACUGGAAGACUGCAG-3'; Oligo 2, 5'-CUGGUGCAGCACUCUCGAC GGCACCUAUCUGCCAUCGUAG-3'; Oligo 3, 5'-CGAUCGUCACUAUCUG CAUCAGUAGCUCUAGCAUACUGAG-3'; Oligo 4, 5'-UCUUUCGUCAGGG CAGAGGAUGGAUGCAAGGAUAAGUGGA-3'. Polyadenylation was performed by using a poly(A) tailing kit (Ambion). 3'-deoxyATP (cordycepin triphosphate, Jena Biosciences) was introduced 10 min after the initiation of the polyadenylation reaction for 3'-end blocking and tail length limitation. Reaction products were cleaned by phenol-chloroform extraction and ethanol precipitation. Samples were analysed with microcapillary electrophoresis (Agilent Technologies) (Supplementary Fig. 1). Poly(A)⁺ *S. cerevisiae* RNA strain DBY746 (*his3D1 leu2-3 leu2-112 ura3-52 trp1-289*), grown under standard conditions (yeast peptone dextrose, 30 °C) was obtained from Clontech (product number 636312). *S. cerevisiae* poly(A)⁺ RNA (2 ng) was used for 3'-end blocking reaction with poly(A) polymerase and 3'-deoxyATP.

Surfaces and template capture. Fifty-nucleotide poly(dT) primers (obtained from IDT) were covalently coupled to sequencing surfaces prepared on glass coverslips in one-channel or five-channel formats. Slides are available from Helicos BioSciences. Poly(A) tail containing RNA molecules were hybridized to the surface at 10–30 pM, requiring 0.5–1.5 fmol polyadenylated RNA per sequencing reaction. The surface was rinsed and the locations of the hybridized templates were determined by imaging after the 'fill and lock' step.

cDNA preparation, PCR amplification and DNA sequencing with the Helicos Genetic Analysis System. First-strand cDNA was prepared using a SuperScript

III first-strand cDNA synthesis kit (Invitrogen) from 500 ng poly(A)⁺ *S. cerevisiae* RNA according to manufacturer's instructions using 50 pmol dT/U-25-V primer (Supplementary Table 1). After cDNA synthesis, RNA was removed by RNase H (Invitrogen) and RNA If (New England Biolabs) digestion for 30 min at 37 °C followed by cleaning with a Nucleotide Removal Kit (Qiagen, 28304). The cDNA is then PCR-amplified with the dT/U-25-V primer and gene/snoRNA specific primers (Supplementary Table 1) using Taq polymerase (New England Biolabs, M0273) under the following thermal cycling conditions: 94 °C for 3 min, 30 cycles of 94 °C for 30 s, 48 °C for 30 s, 72 °C for 30 s, followed by a final 72 °C 10 min incubation step. The excess primers were removed by running the PCR products on 1% agarose gel. Because 3' ends of genes are amplified, and multiple and variable size PCR products are expected, we extracted regions from the gels representing 50–500 base pairs (bp) size distribution (visible PCR products had 100–300 bp sizes) and isolated the DNA with a QIAEX II gel extraction kit (Qiagen, 20021). We chose this approach over commercial column/bead-based cleaning methods because: (1) PCR primers need to be removed as much as possible, otherwise they will be A-tailed by the terminal transferase (described below) and sequenced; and (2) as we expected multiple and variably sized PCR products, we did not want to use commercial systems that may have varying efficiencies of removing small fragments (generally <100 bp) and preserving larger DNA fragments. The PCR products were then treated with the USER enzyme (New England Biolabs, M5505) to eliminate/reduce the 5'-T/U tails left on the PCR products after PCR (Supplementary Fig. 11A). This step was performed to prevent potential competition of the 5'-T/U tail with the Poly(dT) primers on sequencing surfaces used for template capture and sequencing initiation. The USER reaction was cleaned with the Qiagen Nucleotide Removal Kit. Ten nanograms from each PCR product was combined and A-tailed with terminal transferase (New England Biolabs). Briefly, pooled PCR products were heat denatured at 95 °C for 5 min in the presence of the supplied 1× reaction buffer and 2.5 mM CoCl₂, followed by rapid snap-cooling on ice. Terminal transferase (40 U; New England Biolabs) and 900 pmol dATP were then added to the denatured DNA in 50 µl final reaction volume, incubated at 37 °C for 1 h, followed by the inactivation of the enzyme at 70 °C for 10 min. The blocking step was performed by adding 300 pmol ddTTP and 4 U of terminal transferase to the heat-denatured A-tailed reaction above, incubating at 37 °C for 1 h, followed by the inactivation of the enzyme at 70 °C for 20 min. After the tailing and blocking steps, the final DNA was then loaded directly into two channels of the 50 channel Helicos Genetic Analysis System without additional cleaning steps. cDNA sequencing data alignment to the yeast genome (October 2003 assembly) was performed using IndexDP³⁰.

30. Lipson, D. *et al.* Quantification of the yeast transcriptome by single-molecule sequencing. *Nature Biotechnol.* 27, 652–658 (2009).

JAK2 phosphorylates histone H3Y41 and excludes HP1 α from chromatin

Mark A. Dawson^{1,2*}, Andrew J. Bannister^{3*}, Berthold Göttgens¹, Samuel D. Foster¹, Till Bartke³, Anthony R. Green^{1,2*} & Tony Kouzarides^{3*}

Activation of Janus kinase 2 (JAK2) by chromosomal translocations or point mutations is a frequent event in haematological malignancies^{1–6}. JAK2 is a non-receptor tyrosine kinase that regulates several cellular processes by inducing cytoplasmic signalling cascades. Here we show that human JAK2 is present in the nucleus of haematopoietic cells and directly phosphorylates Tyr 41 (Y41) on histone H3. Heterochromatin protein 1 α (HP1 α), but not HP1 β , specifically binds to this region of H3 through its chromo-shadow domain. Phosphorylation of H3Y41 by JAK2 prevents this binding. Inhibition of JAK2 activity in human leukaemic cells decreases both the expression of the haematopoietic oncogene *lmo2* and the phosphorylation of H3Y41 at its promoter, while simultaneously increasing the binding of HP1 α at the same site. These results identify a previously unrecognized nuclear role for JAK2 in the phosphorylation of H3Y41 and reveal a direct mechanistic link between two genes, *jak2* and *lmo2*, involved in normal haematopoiesis and leukaemia^{1–9}.

JAK2 signalling is implicated in various biological processes, including cell cycle progression, apoptosis, mitotic recombination, genetic instability and alteration of heterochromatin^{10–13}. The most common somatic alteration of JAK2 is a gain-of-function mutation (JAK2 V617F) associated with human myeloproliferative diseases^{1,2}. The diverse roles of JAK2 in normal and leukaemic haematopoiesis are believed to be mediated by cytoplasmic signalling pathways^{1,2}.

We found that JAK2 has a previously unrecognized nuclear pool in haematopoietic cells. Figure 1a and Supplementary Fig. 1 show JAK2 within the nuclei of three cell lines (HEL, UKE1 and SET2) harbouring JAK2 V617F¹⁴. However, JAK2 staining was also nuclear in K562 cells, which express wild-type JAK2. Nuclear JAK2 was also observed in primary cells, positive for the CD34 stem-cell antigen, obtained from a patient with JAK2 V617F-positive post-polycythaemic myelofibrosis. Transfection of JAK2 into a JAK2-null background, γ -2A cells¹⁵, independently confirmed the nuclear localization of JAK2 and validated the specificity of the antibodies used in immunofluorescence (Fig. 1b and Supplementary Fig. 2). Finally, subcellular fractionation experiments using HEL cells also demonstrated JAK2 in the nucleus (Fig. 1c). Taken together, these results demonstrate that a significant proportion of JAK2 is present within the nuclei of haematopoietic cells, irrespective of JAK2 mutation status.

To explore the role of JAK2 within the nucleus we investigated the possibility that histones could be a substrate. Among all core histones, we found that recombinant JAK2 specifically phosphorylated histone H3, a reaction inhibited by the JAK2 inhibitor TG101209 (ref. 16) (Fig. 2a and Supplementary Fig. 3). H3 contains three highly conserved tyrosine residues, one of which, H3Y41, is positioned at the amino terminus of the first helix of H3 (the α N1-helix) where

DNA enters the nucleosome (Fig. 2b). Given its accessible location, we reasoned that H3Y41 might be the JAK2 target; we therefore generated an antibody against phosphorylated H3Y41 (H3Y41ph), verified its specificity (Supplementary Figs 4 and 5) and used it to demonstrate

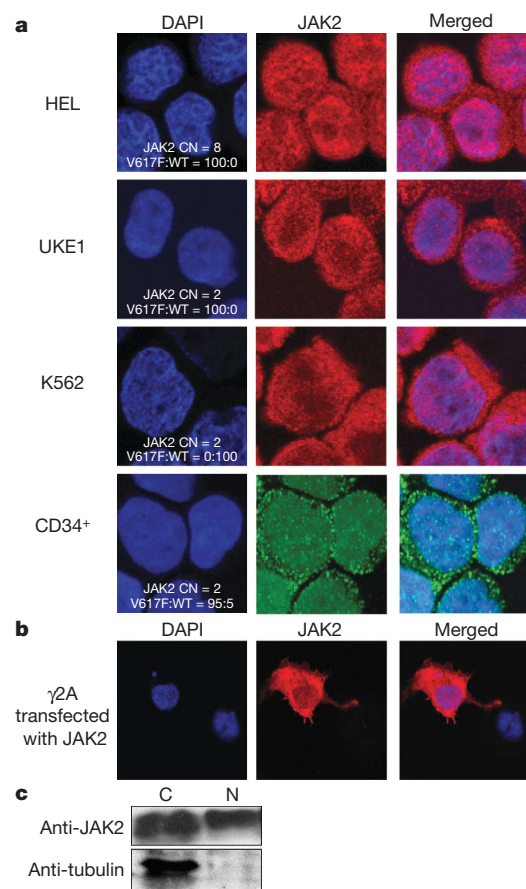


Figure 1 | JAK2 is present in the nucleus of haematopoietic cells. **a**, Confocal immunofluorescent images identify nuclear JAK2 in haematopoietic cell lines and primary CD34⁺ peripheral blood stem cells (CD34⁺). CN, copy number; DAPI, 4,6-diamidino-2-phenylindole; V617F:WT, the ratio of JAK2 V617F to JAK2 wild-type. Two primary anti-JAK2 antibodies were used (detailed in Methods and Supplementary Fig. 2c). **b**, Confocal images of JAK2-null (γ 2A) cells transfected with JAK2. **c**, Western blotting of cytoplasmic (C) and nuclear (N) extracts demonstrates that JAK2 is present in both cellular compartments; however, β -tubulin (anti-tubulin) is confined to the cytoplasmic fraction.

¹Cambridge Institute for Medical Research and Department of Haematology, University of Cambridge, Hills Road, Cambridge CB2 0XY, UK. ²Department of Haematology, Addenbrooke's Hospital, Hills Road, Cambridge CB2 0QQ, UK. ³Gurdon Institute and Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QN, UK. *These authors contributed equally to this work.

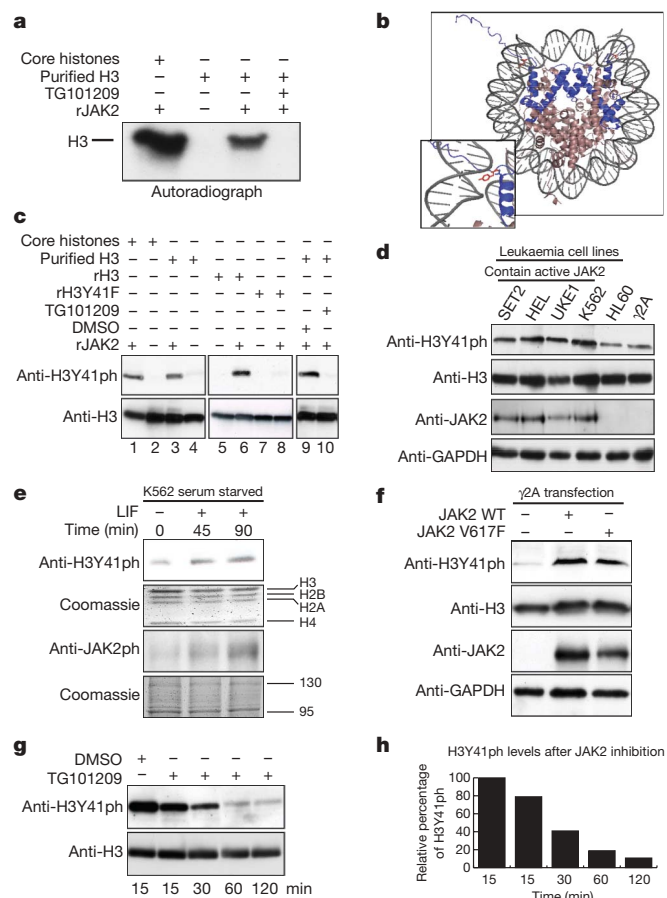


Figure 2 | JAK2 phosphorylates H3Y41 *in vitro* and *in vivo*. **a**, *In vitro* kinase assay using [γ - 32 P]ATP and recombinant JAK2 (rJAK2). **b**, H3 (blue) is shown in the nucleosome; the inset highlights H3Y41 at the amino terminus of the first helix. **c**, *In vitro* kinase assay followed by western blot analysis using the H3Y41ph antibody. **d**, H3Y41ph was measured in chromatin from the indicated cell lines by western blotting. **e**, After serum starvation, K562 cells were stimulated with LIF. H3Y41ph and phospho-JAK2 (anti-JAK2ph) were measured by western blot analyses of whole cell extracts. **f**, H3Y41ph was determined in chromatin from γ 2A cells, transfected with either wild-type JAK2, JAK2 V617F or empty vector. **g**, H3Y41ph was measured in chromatin from HEL cells grown in the presence of a specific JAK2 inhibitor (TG101209) or vehicle control (dimethylsulphoxide; DMSO). **h**, Quantification of western blot in **g**. Similar results were obtained with a second specific JAK2 inhibitor (AT9283; data not shown).

phosphorylation of H3Y41 by recombinant JAK2 *in vitro* (Fig. 2c, lanes 1–6). H3Y41 phosphorylation was inhibited by TG101209 and by mutation of H3Y41 to phenylalanine (Fig. 2c, lanes 7–10). Cellular JAK2, immunoprecipitated from HEL cells, also phosphorylated H3Y41 (Supplementary Fig. 6).

To assess the phosphorylation of H3Y41 *in vivo*, chromatin preparations from six cell lines were probed with the H3Y41ph antibody. H3Y41 phosphorylation was more abundant in cell lines containing active JAK2 (SET2, HEL, UKE1 and K562)¹⁴, whereas H3Y41ph was significantly decreased in HL60 cells and γ 2A cells, which lack detectable JAK2 (refs 15, 17) (Fig. 2d). Stimulation of K562 cells with leukaemia inhibitory factor (LIF) or platelet-derived growth factor-BB (PDGF-BB) resulted in activation of JAK2, as demonstrated by JAK2 phosphorylation, and produced a concomitant increase in H3Y41ph (Fig. 2e and Supplementary Fig. 7a). Moreover, stimulation of murine BaF3 cells with interleukin-3 (IL-3), a cytokine that signals by means of JAK2 in these cells, also increased H3Y41ph (Supplementary Fig. 7b). Taken together, these data demonstrate that H3Y41ph is present *in vivo* and that cytokine signalling regulates its levels.

The presence of residual H3Y41ph in HL60 and JAK2-null γ 2A cells indicates that JAK2 is not the only kinase responsible for this modification. However, transfection of JAK2 into γ 2A cells demonstrated that it is one of the cellular kinases responsible for H3Y41ph *in vivo* (Fig. 2f). To provide further evidence that JAK2 phosphorylates H3Y41 *in vivo*, we used two specific, chemically distinct JAK2 inhibitors, TG101209 and AT9283 (refs 16, 18). Chromatin prepared from HEL cells grown in the presence of either JAK2 inhibitor contained significantly decreased H3Y41ph compared with non-treated cells (Supplementary Fig. 8b). These changes were not a consequence of broad effects on cell cycle or apoptosis (Supplementary Fig. 8a). Moreover, inhibition of JAK2 produced a rapid and sustained loss of H3Y41ph. The decrease in H3Y41ph occurred within 15 min, and by 1 h an 80% decrease was observed (Fig. 2g, h). The rapidity of this response, together with the *in vitro* data, indicates that JAK2 directly phosphorylates H3Y41 *in vivo*.

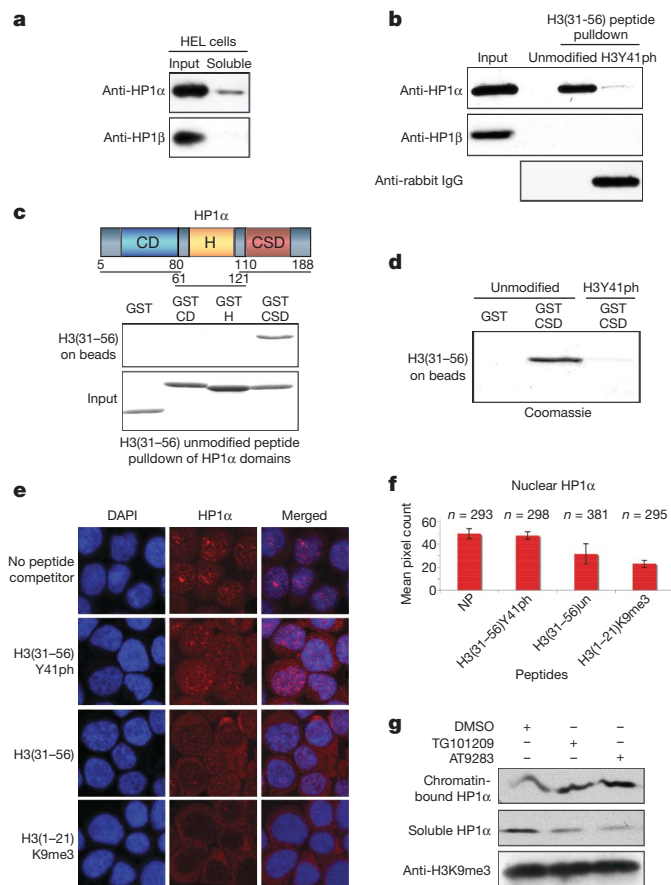


Figure 3 | HP1α binds the Y41 region of H3 in a phosphorylation-dependent manner. **a**, The soluble fraction from permeabilized nuclei was analysed by western blotting for HP1α and HP1β. **b**, HP1α and HP1β were tested for their ability to bind either unmodified H3(31–56) peptides (unmod) or identical peptides phosphorylated at H3Y41 (H3Y41ph). Anti-rabbit IgG detects H3Y41ph antibody. **c**, HP1α contains three domains: a chromo domain (CD), a hinge (H) and a chromo-shadow domain (CSD). The indicated regions were expressed and tested for binding to the unmodified region of H3(31–56). GST, glutathione S-transferase. **d**, CSD binding to H3(31–56) is abrogated by phosphorylation of H3Y41. **e**, HEL60 cells were permeabilized and the binding of HP1α to chromatin was challenged by competition with the indicated peptides. HP1α localization was then revealed by immunofluorescence. **f**, Quantification of the amount of HP1α observed in **e**. H3(31–56)un, unmodified H3(31–56) peptide; n, number of nuclei counted; error bars represent s.d. **g**, HEL cells were treated with DMSO, TG101209 or AT9283. Permeabilized nuclei were prepared and challenged with H3K9me3 peptide to disassociate a relatively small percentage of HP1α from chromatin. Chromatin and soluble fractions were then western blotted for HP1α and H3K9me3.

The JAK pathway in *Drosophila melanogaster* has recently been implicated in the alteration of heterochromatin through the disruption of HP1 (ref. 13). We therefore investigated whether JAK2 signalling in a haematopoietic cell line (HEL) affected the association of HP1 α or HP1 β with chromatin. Figure 3a shows that there was a significant amount of soluble, non-chromatin-bound HP1 α in permeabilized HEL nuclei, whereas HP1 β was essentially bound to chromatin (Fig. 3a). This observation raised the possibility that JAK2 signalling in HEL cells may weaken HP1 α binding and/or stabilize the binding of HP1 β .

Given that JAK2 directly phosphorylates H3Y41, we considered the possibility that H3 may contain an additional binding site for HP1 α or HP1 β near Y41. Figure 3b and Supplementary Fig. 9a, b show that HP1 α bound specifically to an unmodified H3 peptide encompassing amino-acid residues 31–56. HP1 α binding was markedly decreased when the peptide was phosphorylated at Y41. In contrast, HP1 β bound neither the unmodified nor the modified peptide. The integrity of the H3Y41ph peptide was demonstrated by the fact that the H3Y41ph antibody bound only the phosphorylated peptide. Binding to the Y41 region of H3 was specifically mediated by the chromo-shadow domain (CSD) of HP1 α , whereas its chromo domain has been shown to be responsible for its interaction with methylated H3K9 (H3K9me) (Fig. 3c)^{19,20}. Indeed, H3K9me peptides added *in trans* neither stimulate nor inhibit the binding of HP1 α to the Y41

region of H3 (Supplementary Fig. 9c), but H3Y41 phosphorylation inhibited binding of the CSD to H3 (Fig. 3d). Together these data demonstrate that the CSD of HP1 α binds the Y41 region of H3 and that this binding is inhibited by H3Y41 phosphorylation.

To further characterize HP1 α binding to H3 in a more physiological context, we performed immunofluorescence experiments with peptide competition. Figure 3e, f shows that peptides spanning H3 residues 31–56, and peptides containing trimethylated H3K9 (H3K9me3), displaced HP1 α from nuclear heterochromatic speckles¹⁹. In contrast, a H3(31–56) peptide phosphorylated at Y41 (H3Y41ph) was unable to displace HP1 α efficiently from heterochromatin. We next examined whether binding of HP1 α is modulated by JAK2 signalling *in vivo*. Permeabilized nuclei were prepared from HEL cells cultured with or without JAK2 inhibitors. Inhibition of JAK2 increased the proportion of chromatin-bound HP1 α within these nuclei (Fig. 3g; compare lanes 2 and 3 with lane 1). Whereas the level of H3Y41ph was decreased by inhibition of JAK2 (Fig. 2g and Supplementary Fig. 8b), the level of H3K9me3 was unaltered, which is consistent with the concept that JAK2 signalling reduces HP1 α binding by phosphorylating H3Y41.

To investigate the biological consequences of H3Y41 phosphorylation, we used expression arrays to identify JAK2-regulated genes in HEL cells (Fig. 4a and Supplementary Fig. 10a). Of those genes whose messenger RNA levels were most decreased by inhibition of JAK2, several have previously been identified as transcriptional targets of

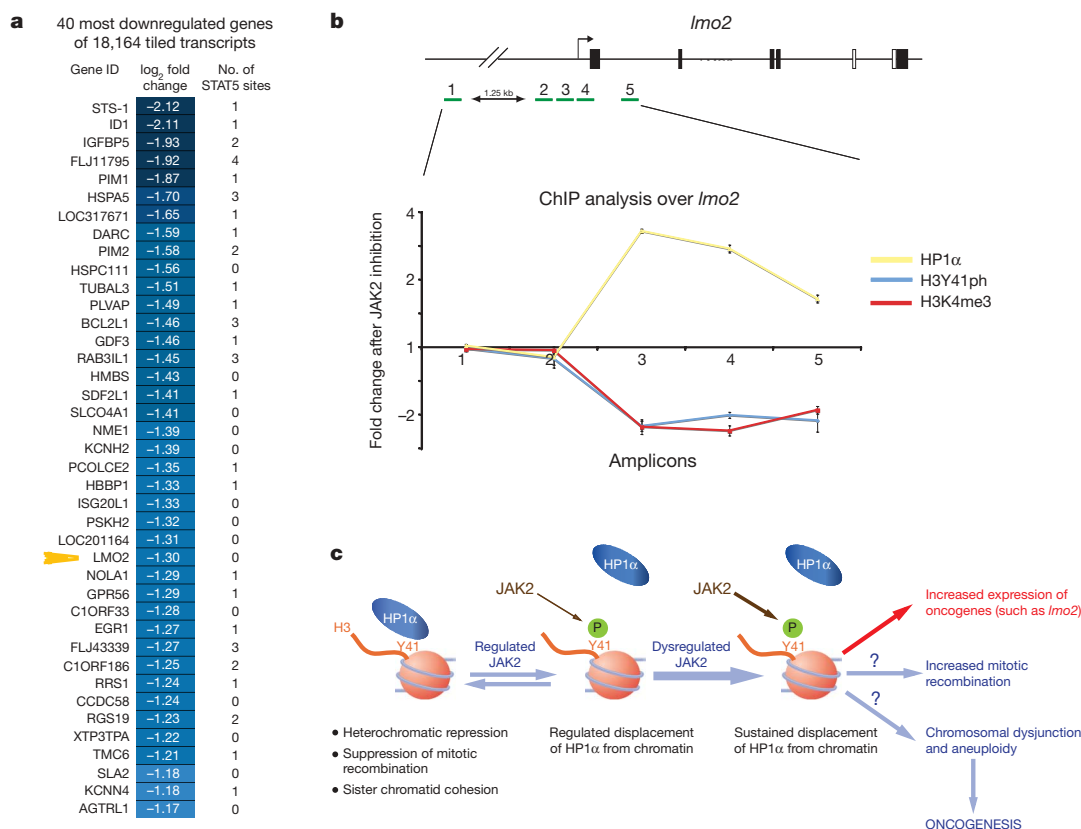


Figure 4 | JAK2 signalling regulates the expression of the *lmo2* oncogene. **a**, Messenger RNA and chromatin (used in **b**) was isolated from HEL cells treated for 4 h with TG101209 or DMSO. Messenger RNA from two biological replicates was used to generate a gene expression profile. The 40 most downregulated genes are illustrated. The number of predicted STAT5 DNA-binding sites in each locus is indicated. The yellow wedge highlights the *lmo2* gene. **b**, Five regions within the *lmo2* locus were investigated (amplicons 1–5; see schematic representation of *lmo2* locus) by chromatin

immunoprecipitation analyses with antibodies against HP1 α , H3Y41ph and H3K4me3. The data were normalized for H3 occupancy. Error bars represent s.d. for each amplicon. **c**, Model depicting the decrease in HP1 α binding to chromatin after phosphorylation of H3Y41 by JAK2. On the left are the known functions of HP1 α ; on the right are the known consequences of dysregulated JAK2 seen as a feature in JAK2-mediated haematological malignancies.

the canonical JAK2–STAT5 pathway (reviewed in ref. 21). However, this approach also identified JAK2-regulated genes, including *lmo2*, that lacked a predicted STAT5-binding site as defined previously²². *lmo2* is essential for normal haematopoietic development, has been implicated in leukaemogenesis^{7,9} and was in the top 0.5% of genes downregulated by inhibition of JAK2. The link between *lmo2* expression and JAK2 inhibition has been noted previously²³; this was further confirmed by quantitative polymerase chain reaction with reverse transcription (RT–PCR) and with a second JAK2 inhibitor, AT9283 (Supplementary Fig. 10b, c). Chromatin immunoprecipitation was then employed to investigate chromatin structure at *lmo2* after inhibition of JAK2. Downregulation of *lmo2* expression (corroborated by decreased levels of H3K4me3) was accompanied by decreased levels of H3Y41ph together with a reciprocal increase in the binding of HP1 α , but not HP1 β , at sites surrounding the *lmo2* transcriptional start site (Fig. 4b, Supplementary Fig. 11 and data not shown). The promoter of *B2M*, a housekeeping gene encoding β_2 -microglobulin, and two sites upstream of the *lmo2* promoter showed no changes in H3K4me3, H3Y41ph or HP1 α (Fig. 4b and Supplementary Fig. 12). Collectively, these results demonstrate that JAK2 signalling results in H3Y41 phosphorylation and the exclusion of HP1 α from the *lmo2* promoter.

The data presented here demonstrate a novel nuclear function for JAK2 that is distinct from its established role as an initiator of cytoplasmic signalling cascades. In the nucleus, JAK2 mediates the phosphorylation of H3Y41 and excludes HP1 α from a new binding site surrounding H3Y41. Given that H3Y41 lies within a region known to affect nucleosome remodelling²⁴, phosphorylation of H3Y41 may regulate chromatin architecture around specific gene promoters.

The displacement of HP1 α by JAK2 is likely to be tightly regulated in normal cells, whereas in malignancies driven by constitutive activation of JAK2, unregulated displacement of chromatin-bound HP1 α may override its potential tumour suppressive functions (Fig. 4c). HP1 α is recognized to reduce mitotic recombination²⁵, repress the transcription of heterochromatic genes²⁶ and preserve centromeric architecture, leading to the faithful segregation of sister chromatids²⁷. Indeed, the phenotypic consequences of constitutive JAK2 activation in haematological malignancies (increased gene expression, mitotic recombination and genetic instability)^{1,2,12} are consistent with the reversal of these HP1 α functions. This suggestion is further supported by the fact that enforced overexpression of HP1 ameliorates the leukaemic phenotype of overactive JAK signalling in *D. melanogaster*¹³.

METHODS SUMMARY

Cell culture and isolation of peripheral blood stem cells were performed with standard methodology¹². Immunofluorescence images were captured with an Olympus Fluoview FV1000 microscope, and cells were prepared and stained as described previously^{12,19}. Cell fractionation, immunoprecipitation, western blotting and kinase assays were performed with standard methodology¹⁹. Peptides (Supplementary Table 1) were synthesized by Almac Sciences and used for binding/competition assays as described previously¹⁹.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 27 April; accepted 21 August 2009.

Published online 27 September 2009.

1. Campbell, P. J. & Green, A. R. The myeloproliferative disorders. *N. Engl. J. Med.* **355**, 2452–2466 (2006).
2. Levine, R. L., Pardanani, A., Tefferi, A. & Gilliland, D. G. Role of JAK2 in the pathogenesis and therapy of myeloproliferative disorders. *Nature Rev. Cancer* **7**, 673–683 (2007).
3. Scott, L. M. *et al.* JAK2 exon 12 mutations in polycythemia vera and idiopathic erythrocytosis. *N. Engl. J. Med.* **356**, 459–468 (2007).
4. Lacronique, V. *et al.* A TEL–JAK2 fusion protein with constitutive kinase activity in human leukemia. *Science* **278**, 1309–1312 (1997).
5. Bericovich, D. *et al.* Mutations of JAK2 in acute lymphoblastic leukaemias associated with Down's syndrome. *Lancet* **372**, 1484–1492 (2008).

6. Mullighan, C. G. *et al.* JAK mutations in high-risk childhood acute lymphoblastic leukemia. *Proc. Natl Acad. Sci. USA* **106**, 9414–9418 (2009).
7. McCormack, M. P. & Rabbitts, T. H. Activation of the T-cell oncogene *LMO2* after gene therapy for X-linked severe combined immunodeficiency. *N. Engl. J. Med.* **350**, 913–922 (2004).
8. Neubauer, H. *et al.* Jak2 deficiency defines an essential developmental checkpoint in definitive hematopoiesis. *Cell* **93**, 397–409 (1998).
9. Yamada, Y. *et al.* The T cell leukemia LIM protein Lmo2 is necessary for adult mouse hematopoiesis. *Proc. Natl Acad. Sci. USA* **95**, 3890–3895 (1998).
10. Walz, C. *et al.* Activated Jak2 with the V617F point mutation promotes G1/S phase transition. *J. Biol. Chem.* **281**, 18177–18183 (2006).
11. Zhao, R. *et al.* Inhibition of the Bcl-xL deamidation pathway in myeloproliferative disorders. *N. Engl. J. Med.* **359**, 2778–2789 (2008).
12. Plo, I. *et al.* JAK2 stimulates homologous recombination and genetic instability: potential implication in the heterogeneity of myeloproliferative disorders. *Blood* **112**, 1402–1412 (2008).
13. Shi, S. *et al.* JAK signaling globally counteracts heterochromatic gene silencing. *Nature Genet.* **38**, 1071–1076 (2006).
14. Quantmeier, H., MacLeod, R. A., Zaborski, M. & Drexler, H. G. JAK2 V617F tyrosine kinase mutation in cell lines derived from myeloproliferative disorders. *Leukemia* **20**, 471–476 (2006).
15. Watling, D. *et al.* Complementation by the protein tyrosine kinase JAK2 of a mutant cell line defective in the interferon- γ signal transduction pathway. *Nature* **366**, 166–170 (1993).
16. Pardanani, A. *et al.* TG101209, a small molecule JAK2-selective kinase inhibitor potentially inhibits myeloproliferative disorder-associated JAK2V617F and MPLW515L/K mutations. *Leukemia* **21**, 1658–1668 (2007).
17. Xie, S. *et al.* Involvement of Jak2 tyrosine phosphorylation in Bcr–Abl transformation. *Oncogene* **20**, 6188–6195 (2001).
18. Squires, M. S. *et al.* AT9283, a potent inhibitor of JAK2, is active in JAK2 V617F myeloproliferative disease models. *Blood* **110**, Abstract 3537 (2007).
19. Bannister, A. J. *et al.* Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **410**, 120–124 (2001).
20. Lachner, M., O'Carroll, D., Rea, S., Mechtler, K. & Jenuwein, T. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **410**, 116–120 (2001).
21. Wood, A. D. *et al.* Id1 promotes expansion and survival of primary erythroid cells and is a target of JAK2V617F–STAT5 signalling. *Blood* **114**, 1820–1830 (2009).
22. Donaldson, I. J., Chapman, M. & Gottgens, B. TFBScluster: a resource for the characterization of transcriptional regulatory networks. *Bioinformatics* **21**, 3058–3059 (2005).
23. Ma, A. C., Ward, A. C., Liang, R. & Leung, A. Y. The role of jak2a in zebrafish hematopoiesis. *Blood* **110**, 1824–1830 (2007).
24. Ferreira, H., Somers, J., Webster, R., Flaus, A. & Owen-Hughes, T. Histone tails and the H3 α N helix regulate nucleosome mobility and stability. *Mol. Cell. Biol.* **27**, 4037–4048 (2007).
25. Cummings, W. J. *et al.* Chromatin structure regulates gene conversion. *PLoS Biol.* **5**, e246 (2007).
26. Pantaleeva, I. *et al.* HP1 α guides neuronal fate by timing E2F-targeted genes silencing during terminal differentiation. *EMBO J.* **26**, 3616–3628 (2007).
27. Yamagishi, Y., Sakuno, T., Shimura, M. & Watanabe, Y. Heterochromatin links to centromeric protection by recruiting shugoshin. *Nature* **455**, 251–255 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank P. Flicek, S. Wilder, B. Huntly, S. J. Dawson and all the members of the A.R.G., B.G. and T.K. laboratories, in particular P. Hurd, B. Xhemalce, E. J. Baxter and P. Beer, for helpful discussions; A. Wood for sharing unpublished data; and J. LeQuesne for help with image analysis. This work was supported by PhD fellowship grants to M.A.D. from the General Sir John Monash Foundation, the Cambridge Commonwealth Trust and Raymond and Beverly Sackler. The Green (A.R.G.) laboratory is funded by the UK Leukaemia Research Fund, the Wellcome Trust, the Leukemia & Lymphoma Society of America and the National Institute for Health Research Cambridge Biomedical Research Centre. The Gottgens (B.G.) laboratory is funded by the Leukaemia Research Fund, Cancer Research UK, the Leukemia & Lymphoma Society of America and a Medical Research Council studentship to S.D.F. The Kouzarides (T.K.) laboratory is funded by grants from Cancer Research UK and the 6th Research Framework Programme of the European Union (Epitron, HEROIC and SMARTER).

Author Contributions M.A.D. and A.J.B. designed experiments, performed research, interpreted data and wrote the manuscript. S.D.F. and T.B. performed experiments. B.G., A.R.G. and T.K. designed experiments, interpreted data and wrote the manuscript. M.A.D. and A.J.B. are joint first authors. A.R.G. and T.K. are joint senior authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.R.G. (arg1000@cam.ac.uk) or T.K. (t.kouzarides@gurdon.cam.ac.uk).

METHODS

Cell culture and transfection. HEL, SET2, UKE-1, HL60, K562 and BaF3 were grown in RPMI 1640 medium (Sigma-Aldrich), and γ 2A cells were grown in DMEM medium (Sigma-Aldrich). All growth media were supplemented with 10% fetal calf serum and 1% penicillin/streptomycin. BaF3 cells were grown in the presence of 10 ng ml⁻¹ IL-3. Cells were incubated at 37 °C and 5% CO₂. Transient transfection of γ 2A cells was performed with FuGENE (Roche Applied Science) in accordance with the manufacturer's instructions. Cytokine stimulation was performed in K562 and BaF3 cells after 72 h of serum starvation. LIF (1,000 IU ml⁻¹), PDGF-BB (10 ng ml⁻¹) and IL-3 (10 ng ml⁻¹) were used individually to stimulate the cells for up to 90 min. Mouse embryonic fibroblasts with the functional status of H2AX WT (*H2AX*^{+/+}) and H2AX-null (*H2AX*^{-/-}) were provided by K. Miller and S. P. Jackson.

Isolation of peripheral blood stem cells. Mononuclear cells from peripheral blood were separated over a Ficoll density gradient. CD34⁺ cells were then purified by a double-positive magnetic cell sorting system (AutoMACS; Miltenyi Biotec), in accordance with the manufacturer's instructions.

Immunofluorescence microscopy. HEL, SET2, UKE-1, HL60, K562 and CD34⁺ cells were washed once in 1 × PBS before cytocentrifugation onto polylysine-coated microscope slides. γ 2A cells were grown on coverslips before washing in 1 × PBS. Cells were fixed for 30 min in methanol at -20 °C. After stepwise incubation with a primary antibody, and then a secondary fluorescent antibody, cells were stained with Hoechst 33258 (Sigma-Aldrich) and mounted with Vectashield mounting medium (Vector laboratories). Confocal laser images were captured with an Olympus Fluoview FV1000 microscope equipped with a 40× oil-immersion lens. Image processing was performed with Photoshop (Adobe Systems).

Cell fractionation, immunoprecipitation and immunoblotting. Cytoplasmic, nucleosolic and chromatin fractions were prepared from cells as described previously²⁸. In brief, cells were washed twice in 1 × PBS and once in buffer A (10 mM HEPES pH 7.9, 1.5 mM MgCl₂, 10 mM KCl, 0.5 mM dithiothreitol (DTT) and protease inhibitor cocktail). Cells were then pelleted and resuspended in buffer A with 0.1% (v/v) Nonidet P40 and incubated on ice for 10 min. The supernatant containing the cytoplasmic fraction was collected after centrifugation and the pellet was resuspended in an equal volume (relative to the cytoplasmic extract) of buffer B (20 mM HEPES pH 7.9, 1.5 mM MgCl₂, 300 mM NaCl, 0.5 mM DTT, 25% (v/v) glycerol, 0.25% Triton X-100, 0.2 mM EDTA and protease inhibitor cocktail). After centrifugation, the supernatant contains the nucleosolic fraction and the insoluble pellet is composed primarily of chromatin and associated proteins. Equal volumes of cytoplasmic and nucleosolic fractions were separated by SDS-PAGE, transferred to nitrocellulose and probed with relevant antibodies. For immunoprecipitation, cells were lysed in IPH (150 mM NaCl, 50 mM Tris-HCl pH 8.0, 5 mM EDTA, 0.5% (v/v) Nonidet P40) on ice for 15 min and the supernatant was collected after centrifugation and used for immunoprecipitation. Sodium orthovanadate (1 mM) was added to all solutions when performing assays relating to the study of tyrosine phosphorylation. Extracted proteins were mixed with 2 × Laemmli sample buffer, separated by SDS-PAGE, transferred to nitrocellulose or poly(vinylidene difluoride) (PVDF) membranes (Millipore) and stained with Ponceau S to ensure equal transfer. Membranes were then sequentially incubated with primary antibodies and secondary antibodies conjugated with horseradish peroxidase. Membranes were then incubated for enhanced chemiluminescence (ECL[®]; GE Healthcare) and proteins were detected by exposure to X-ray film. Dot-blot assays were performed by spotting synthetic peptide onto pre-wetted PVDF membrane. The membrane was then sequentially probed with primary and secondary antibodies as above. If appropriate, the primary antibody incubation was performed in the presence of competitor peptides (1.0 μ g ml⁻¹) as indicated in the relevant figure panel.

Kinase assays. The JAK2 Enzymatic Assay Kit, HTScan (Cell Signaling Technology) containing active JAK2 as a glutathione S-transferase (GST) fusion protein, recombinant AKT1 (Cell Signaling Technology) and human JAK2 immunoprecipitated (see antibodies) from HEL cells were used in kinase assays *in vitro* employing the same reaction conditions. In brief, assays were performed in 50 μ l of kinase buffer (60 mM HEPES pH 7.9, 5 mM MgCl₂, 5 mM MnCl₂, 3 μ M Na₂VO₄, 1.25 mM DTT, 20 μ M ATP). [γ -³²P]ATP (370 kBq; 222 TBq mmol⁻¹; Perkin Elmer) was added to the buffer in the radiolabelled kinase assays. Calf thymus histones (2 μ g) (core histone, an equimolar mixture of H3, H2A, H2B and H4, 10223565001, Roche; and purified H3 from calf thymus, 11034758001; Roche) or recombinant histones were used as substrates. Recombinant histones were expressed in, and purified from, bacteria as described previously²⁹.

Site-directed mutagenesis. Mutagenesis to introduce the H3Y41F mutant into human histone H3 was performed with the Quickchange Site-directed Mutagenesis kit (Stratagene) in accordance with the manufacturer's instructions.

JAK2 inhibitors. AT9283 was provided by J. Lyons and M. Squires. TG101209 (TargeGen Inc.) is commercially available and was used at 10 nM in the *in vitro* kinase assays and at 3 μ M *in vivo*. AT9283 was used at 300 nM *in vivo*.

Antibodies. The principal H3Y41ph antibody used in the manuscript was anti-H3Y41ph (ab26310; Abcam), and it was used for western blotting at 1:1,000 dilution; two further H3Y41ph antibodies were raised in rabbits against a Y41ph peptide spanning residues H3(37–48) by using the Eurogentec 28-day commercial protocol. The following antibodies were also used at the stated dilutions: anti-JAK2 antibodies (D2E12 no. 3230; Cell Signaling Technology), (IMG-3007; Imgenex) western blot 1:1,000; immunofluorescence 1:100, immunoprecipitation 1:250; (AB3804; Millipore) western blot 1:1,000 and phospho-JAK2 (ab32101) western blot 1:500; anti- β -tubulin (T5201; Sigma-Aldrich) western blot 1:750; anti-phosphotyrosine (4G10; Millipore) western blot 1:1,000; anti-H3 (ab1791; Abcam) western blot 1:10,000; anti-GAPDH (ab9483; Abcam) western blot 1:5,000; anti-H2Ax (ab1175; Abcam) western blot 1:5,000; anti-Flag (Sigma-Aldrich) immunoprecipitation 1:250; anti-HP1 α (no. 2616; Cell Signaling Technology) western blot 1:1,000, immunofluorescence 1:400; anti-HP1 α (clone15.19s2; no. 05-689; Millipore); anti-HP1 β (no. 2613; Cell Signaling Technology) western blot 1:1,000; Texas-red-conjugated IgG (Invitrogen) immunofluorescence 1:250; and Alex Fluor-488-conjugated IgG (Invitrogen) immunofluorescence 1:250.

Recombinant protein production. Recombinant proteins were expressed in and purified from *Escherichia coli* as described³⁰. Mouse full-length HP1 isoforms and the chromo domain (residues 5–80), hinge (residues 61–121) and chromo-shadow domain (residues 110–188) of HP1 α were cloned into pGex vector and expressed as a GST fusion protein.

Pulldown assays. GST fusion proteins and biotin-conjugated peptides were incubated for 1 h with glutathione-agarose beads or streptavidin-Sepharose beads, respectively, in binding buffer (150 mM NaCl, 50 mM Tris-HCl pH 8.0, 5 mM EDTA, 0.25% (v/v) Nonidet P40) at room temperature (between 20 and 25 °C). After being washed three times in binding buffer, the beads were incubated with their potential binding proteins for 1 h at room temperature. The beads were then washed four times with binding buffer, after which bound protein was eluted with hot 2 × Laemmli sample buffer.

Preparation of nuclei for assessing soluble and chromatin-bound HP1 α / β . HEL nuclei were purified³¹ and permeabilized³² as described, except that 300 mM NaCl and 0.25% (v/v) Triton X-100 were used. Nuclei were pelleted and the chromatin fraction was separated from the soluble nuclear fraction as described previously¹. The chromatin and soluble nuclear fractions were then analysed by western blotting for HP1 α and HP1 β . When comparing the localization of HP1 α in the presence or absence of JAK2 inhibitors, a batch of cells was split into three equal amounts and incubated with vehicle alone (DMSO), TG101209 (3 μ M) or AT9283 (300 nM) for 4 h before the isolation of nuclei. Permeabilized nuclei were diluted into PBS and incubated for 2 h with 0.75 ng ml⁻¹ H3K9me3 peptide on ice. They were then processed as above.

Peptide competition and immunofluorescence of mammalian cells. HL60 cells were cytospun onto polylysine-coated slides and fixed for 2 min in ice-cold methanol (containing 10 mg ml⁻¹ peptide, where used). They were then blocked for 15 min in 3% bovine serum albumin, 0.6% (v/v) Triton X-100 in PBS (containing 10 μ g ml⁻¹ peptide, where used). We performed staining with the anti-HP1 α antibody. Antibody incubations contained 10 μ g ml⁻¹ peptide (where used). The displacement of HP1 α from the nucleus was quantified with Image J software (National Institutes of Health).

Gene expression and computational analysis. HEL cells were treated for 4 h with either TG101209 JAK2 inhibitor or DMSO (vehicle) alone. From these cells, mRNA was prepared. Gene expression changes (log₂ fold with and without inhibitor) of duplicate expression profiling samples were calculated with Bioconductor (<http://www.bioconductor.org/>). Using Illumina Gene IDs, we obtained Ensembl gene coordinates for human genome build NCBI 36.1, using Biomart (<http://www.biomart.org/>). To map conserved STAT5-binding sites in non-coding sequences, we generated a genome-wide data set of STAT5 motifs (TTCYNRGAA) conserved in human/mouse whole-genome alignments obtained from the UCSC genome browser (<http://genome.ucsc.edu/>) using the TFBScluster program (http://hsc1.cimr.cam.ac.uk/TFBScluster_genome_portal.html) with the non-exact search parameters (for example, the ambiguous letters YNR may differ between the human and mouse sequences). Finally, the number of conserved STAT5 sites in each gene locus (ranging from 50 kilobases 5' of the first exon to 50 kilobases 3' of the last exon) was calculated with an in-house PERL script.

Chromatin immunoprecipitation assay and real-time PCR analysis. HEL cells were treated for 4 h with either TG101209 JAK2 inhibitor or DMSO (vehicle) alone. From these cells, chromatin was prepared and chromatin immunoprecipitation was performed as described previously³³, with the following important exceptions. Cells were crosslinked with 1% (v/v) formaldehyde for 15 min at room temperature; crosslinking was stopped by the addition of 0.125 M glycine.

The time and percentage of formaldehyde crosslinking is crucial to ensure optimal recognition of the H3Y41ph and HP1 α epitopes. Cells were then lysed in 1% (w/v) SDS, 10 mM EDTA, 50 mM Tris-HCl pH 8.0, 1 mM sodium orthovanadate and protease inhibitors. Cells were sonicated in a Bioruptor (Diagenode) to achieve a mean DNA fragment size of 500 base pairs. An equal volume of Protein-A and Protein-G agarose beads, pre-absorbed with sonicated salmon-sperm DNA and BSA, were used to pre-clear the chromatin for 2 h before immunoprecipitation. Immunoprecipitation was performed for a minimum of 12 h at 4 °C in modified RIPA buffer (1% (v/v) Triton X-100, 0.1% (w/v) sodium deoxycholate, 0.1% (w/v) SDS, 90 mM NaCl, 10 mM Tris-HCl pH 8.0, 1 mM sodium orthovanadate, EDTA-free protease inhibitors). An equal volume of Protein-A and Protein-G agarose beads, pre-absorbed with sonicated salmon-sperm DNA and BSA, were used to bind the antibody and associated chromatin. The beads were washed before elution of the antibody-bound chromatin. Reverse crosslinking of DNA was followed by DNA purification with the QIAquick PCR purification kit (Qiagen). Immunoprecipitated DNA was analysed on an ABI 7300 real-time PCR machine, with power SYBRgreen PCR mastermix in accordance with the manufacturer's instructions. The following primer pairs were used in the chromatin immunoprecipitation analysis: *Lmo2* amplicon 1, 5'-CAGGCTTCTCCCGTGTAAGT-3' (forward) and 5'-AGGACCTCACACGTTGAAGACA-3' (reverse); *Lmo2* amplicon 2, 5'-AGGGAAGTATGACACAATCGAACA-3' (forward) and 5'-TGGCAGAGCCCGTATGCTA-3' (reverse); *Lmo2* amplicon 3, 5'-CCAGACAACTCAAATAACGTACACA-3' (forward) and 5'-AGTGGGTACCATTTGCCCTGTT-3' (reverse); *Lmo2* amplicon 4, 5'-CCTACTCAGAATGTGGAGACTTGTG-3' (forward) and 5'-TGGCCCTCGGAATTGGA-3' (reverse); *Lmo2* amplicon 5, 5'-GGACTTCGCTCT

TCCATCCA-3' (forward) and 5'-GGCATCGGTGTCAGACCAA-3' (reverse); β_2 -microglobulin, 5'-TGGGCACGCGTTTAATATAAGTG-3' (forward) and 5'-GCCCCGAATGCTGTGAGCTT-3' (reverse).

Messenger RNA was prepared from cell extracts with the Qiagen RNeasy kit in accordance with the manufacturer's instructions. Complementary DNA was then prepared with Superscript III Reverse Transcriptase (Invitrogen) and analysed on an ABI 7300 real-time PCR machine, using power SYBRgreen PCR mastermix in accordance with the manufacturer's instructions. The following primer pairs were then used in the cDNA analysis: *Lmo2*, 5'-CGGCGCCTCTACTACAAACT-3' (forward) and 5'-GAATCCGCTTGTACAGGAT-3' (reverse); β_2 -microglobulin, 5'-TGACTTTGTACAGCCCAAG-3' (forward) and 5'-AGCAAGCAAGCAGAATTG-3' (reverse).

28. Osborn, L., Kunkel, S. & Nabel, G. J. Tumor necrosis factor α and interleukin 1 stimulate the human immunodeficiency virus enhancer by activation of the nuclear factor κ B. *Proc. Natl Acad. Sci. USA* **86**, 2336–2340 (1989).
29. Dyer, P. N. *et al.* Reconstitution of nucleosome core particles from recombinant histones and DNA. *Methods Enzymol.* **375**, 23–44 (2004).
30. Bannister, A. J. & Kouzarides, T. The CBP co-activator is a histone acetyltransferase. *Nature* **384**, 641–643 (1996).
31. Krude, T. Mimosine arrests proliferating human cells before onset of DNA replication in a dose-dependent manner. *Exp. Cell Res.* **247**, 148–159 (1999).
32. Krude, T. Initiation of human DNA replication *in vitro* using nuclei from cells arrested at an initiation-competent state. *J. Biol. Chem.* **275**, 13699–13707 (2000).
33. Nielsen, S. J. *et al.* Rb targets histone H3 methylation and HP1 to promoters. *Nature* **412**, 561–565 (2001).

Structural insights into mechanisms of the small RNA methyltransferase HEN1

Ying Huang¹, Lijuan Ji², Qichen Huang¹, Dmitry G. Vassilyev¹, Xuemei Chen² & Jin-Biao Ma^{1,3}

RNA silencing is a conserved regulatory mechanism in fungi, plants and animals that regulates gene expression and defence against viruses and transgenes¹. Small silencing RNAs of ~20–30 nucleotides and their associated effector proteins, the Argonaute family proteins, are the central components in RNA silencing². A subset of small RNAs, such as microRNAs and small interfering RNAs (siRNAs) in plants, Piwi-interacting RNAs in animals and siRNAs in *Drosophila*, requires an additional crucial step for their maturation; that is, 2'-O-methylation on the 3' terminal nucleotide^{3–6}. A conserved S-adenosyl-L-methionine-dependent RNA methyltransferase, HUA ENHANCER 1 (HEN1), and its homologues are responsible for this specific modification^{3–5,7,8}. Here we report the 3.1 Å crystal structure of full-length HEN1 from *Arabidopsis* in complex with a 22-nucleotide small RNA duplex and cofactor product S-adenosyl-L-homocysteine. Highly cooperative recognition of the small RNA substrate by multiple RNA binding domains and the methyltransferase domain in HEN1 measures the length of the RNA duplex and determines the substrate specificity. Metal ion coordination by both 2' and 3' hydroxyls on the 3'-terminal nucleotide and four invariant residues in the active site of the methyltransferase domain suggests a novel Mg²⁺-dependent 2'-O-methylation mechanism.

HEN1 was first identified in a genetic screen as a floral patterning gene and later found to be essential for *Arabidopsis* microRNA (miRNA) accumulation *in vivo*^{9,10}. Subsequently, HEN1 was demonstrated to be a methyltransferase for miRNAs and all types of siRNAs in plants^{3,11}. The 2'-O-methylation protects miRNAs and siRNAs from 3'-end uridylation and 3'-to-5' exonuclease-mediated degradation in *Arabidopsis*^{12,13}. The plant HEN1 and its animal homologues share a highly conserved methyltransferase (MTase) domain¹⁴ (Fig. 1e) that is not closely related to any known RNA 2'-O-MTases according to a phylogenetic analysis¹⁵. Two putative RNA binding modules, a double-stranded RNA binding domain (dsRBD) and a La motif have been identified in the amino-terminal region of HEN1 (ref. 15). To understand the specific recognition of small RNA substrates and the molecular mechanism of the 3'-end 2'-OH-specific methylation by HEN1 and its homologues, we determined the crystal structure of full-length *Arabidopsis* HEN1 in complex with a small RNA duplex in the presence of the cofactor product adenosyl-L-homocysteine (AdoHcy).

The recombinant full-length *Arabidopsis* HEN1 (residue 1–942) was co-crystallized with AdoHcy and a 22-nucleotide small RNA duplex containing a fully complementary 20-nucleotide segment (Fig. 1f) derived from a natural substrate of HEN1, miR173/miR173* (refs 3, 11) (Supplementary Fig. 2c). The crystal structure was determined at 3.1 Å as described in Methods. The structure revealed that *Arabidopsis* HEN1 binds to the small RNA substrate as a monomer (Fig. 1), which is supported by results from gel filtration

experiments (Supplementary Fig. 2). The small RNA substrate exhibits an A-form conformation in the ternary complex structure and both duplex termini are specifically recognized by HEN1. The HEN1 protein consists of five structural domains (Fig. 1e), four of which directly interact with the small RNA substrate (Fig. 1a–c) with the exception of the PPIase-like domain (PLD) which shows a high degree of structural similarity to well characterized FK506-binding proteins¹⁶. The A-form duplex of the small RNA substrate is bound by two double-stranded RNA (dsRNA)-specific binding domains¹⁷, dsRBD1 and dsRBD2. The [5'-m:3'-u] terminus containing the 3'-end 2-nucleotide overhang of the strand that is not methylated (u strand) (Fig. 1f) is bound by the La-motif-containing domain (LCD). Meanwhile, the 3'-end 2-nucleotide overhang of the strand that is methylated (m strand) (Fig. 1f) is deeply buried into the active site of the MTase domain (Fig. 1c). The interface between HEN1 and the small RNA substrate buries a total solvent-accessible surface area of ~5,000 Å² (Fig. 1c), of which dsRBD1, dsRBD2, LCD and the MTase domain each contributes 31%, 13%, 17% and 33%, respectively.

Structure-based sequence alignment and structural superimposition revealed that both dsRBDs contain distinct long insertions in the loop between β1 and β2 (Supplementary Fig. 4). The insertion in dsRBD1 is well defined (Supplementary Fig. 4b), in which a conserved hydrophobic patch stacks over the carboxy-terminal β-strand in the MTase domain (Supplementary Fig. 4c). The insertion in dsRBD2 is longer but less conserved than that in dsRBD1 in most plant HEN1 proteins (Supplementary Fig. 4a) and is completely disordered in the current structure (Supplementary Fig. 4b). Three conserved RNA binding motifs in canonical dsRBDs¹⁷ can be identified in dsRBD1 (Fig. 2a), whereas only two RNA binding motifs are identified in dsRBD2 owing to the disordered loop between β1 and β2 (Fig. 2b). As revealed by buried surface analysis, the interaction of dsRBD1 with the RNA substrate is more extensive than that of dsRBD2. Compared to dsRBD1, dsRBD2 shifts by approximately 3 Å away from the RNA duplex, which may favour binding small RNA duplexes with bulges that are common among miRNAs. The binding of the RNA duplex by dsRBD1 has a key role in substrate recognition, as deletion of dsRBD1 markedly reduced the substrate binding by HEN1, as determined by a cross-linking binding assay (Supplementary Fig. 5a), and its activity, as revealed by a small RNA methyltransferase assay (Supplementary Fig. 5b).

As predicted by bioinformatics analysis¹⁵, the N-terminal half of the LCD contains a La motif fold (Supplementary Fig. 6) that has been shown to specifically bind RNA 3' ends through synergistic cooperation with an RNA recognition motif in the La protein^{18,19}. However, recognition of the [5'-m:3'-u] duplex terminus of the small RNA substrates by the La motif and the C-terminal portion of LCD in HEN1 (Fig. 1f) is different from that observed in the human La protein²⁰. The 3'-terminal nucleotide binding pocket in

¹Department of Biochemistry and Molecular Genetics, Schools of Medicine and Dentistry, University of Alabama at Birmingham, Birmingham, Alabama 35294, USA. ²Department of Botany and Plant Sciences, Institute of Integrative Genome Biology, University of California, Riverside, California 92521, USA. ³Comprehensive Cancer Center, University of Alabama at Birmingham, Birmingham, Alabama 35294, USA.

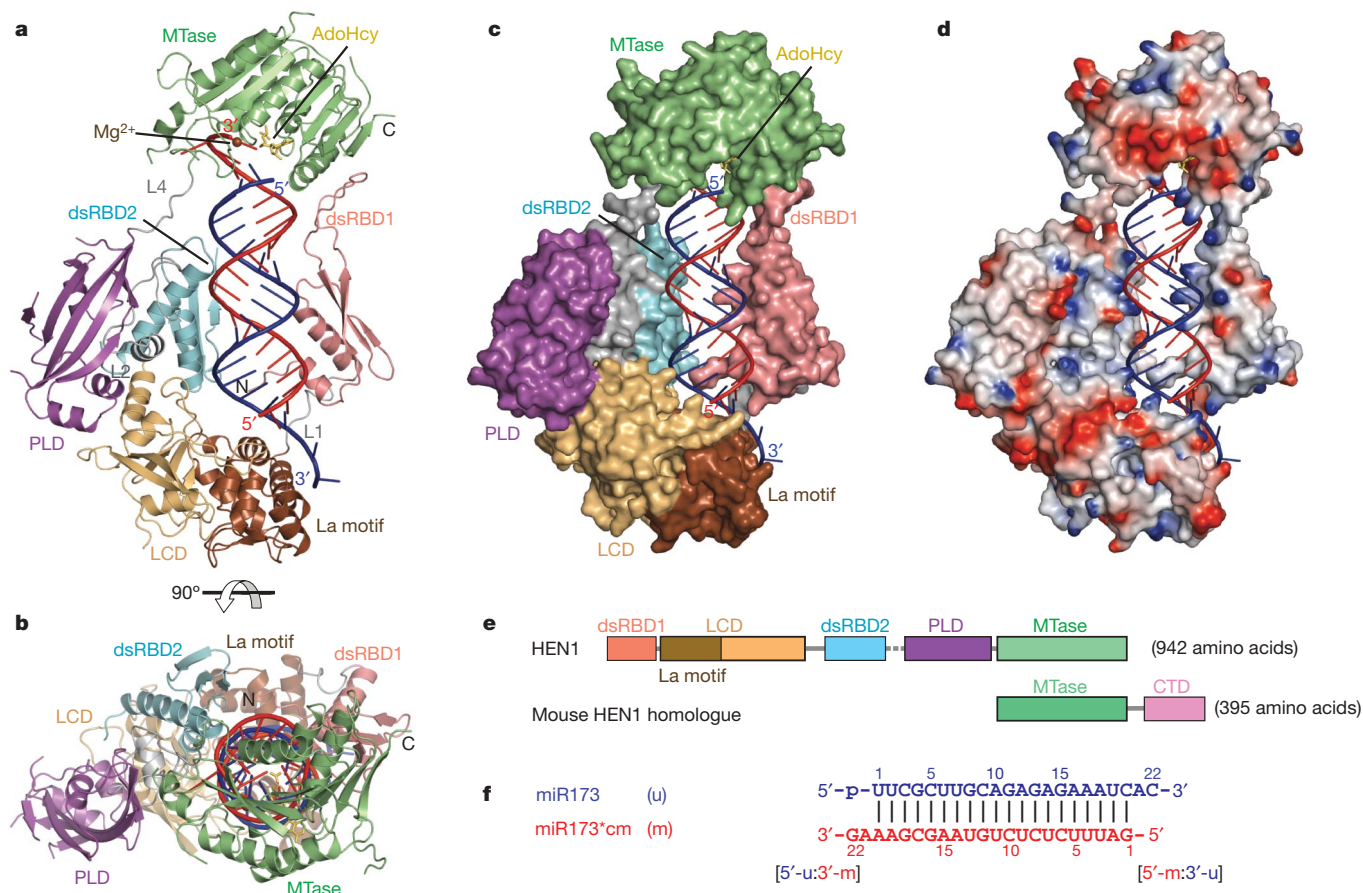


Figure 1 | Structures of HEN1 in complex with a small RNA duplex and AdoHcy. **a**, Ribbon diagram of the complex. dsRBD, violet; La motif, chocolate; LCD, wheat; dsRBD2, cyan; PLD, purple; MTase, green; linkers including L1, L2 and L4, grey; RNA strand to be methylated (m strand), red; RNA strand not to be methylated (u strand), blue; AdoHcy, yellow; Mg^{2+} , brown. **b**, Ribbon diagram of the complex rotated by 90° about the horizontal axis relative to **a**. **c**, **d**, Surface and surface charge views of HEN1

the human La N-terminal domain (NTD)²⁰ is occupied by two conserved residues—H120 and P121—within a HEN1-specific insertion (Supplementary Fig. 6c). The 2-nucleotide 3' overhang of the u strand is looped out from the RNA duplex towards the first α -helix of the La motif and the phosphate of the overhang is bound by Y109 from the first α -helix of the La motif (Fig. 2c). Mutation of Y109 to alanine has no detectable effect on substrate binding or HEN1 activity (Supplementary Fig. 5a, b), indicating that this interaction is not essential for the interaction with small RNA substrates. This result is also consistent with a previous study showing that mutation of the 3'-end 2-nucleotide overhang of the u strand to either a 1-nucleotide or 3-nucleotide overhang has no effect on HEN1 activity on the m strand¹¹.

Furthermore, W333, a conserved residue within a loop from the C-terminal portion of the LCD, stacks over the base of the 5'-terminal nucleotide G1_m (Fig. 2c), and the side chain of W333 occupies the same position as the base of the antepenultimate nucleotide in the structure of the La NTD–RNA complex²⁰ (Supplementary Fig. 6c). Therefore, W333 exactly stacks on H120 and P121 in the La motif of the LCD, which may stabilize the stacking interaction between W333 and the 5'-terminal nucleotide. This end-capping interaction has an essential role in the recognition of small RNA substrates, because the W333A mutant loses both RNA binding ability and small RNA methyltransferase activity (Supplementary Fig. 5a, b). A similar interaction between small RNA duplex and tryptophan residues has been observed in structures of the viral RNA silencing suppressor p19–small

RNA complexes^{21,22}. In fact, p19 interferes with small RNA 3'-end methylation by HEN1 (ref. 23). In addition, recognition of the [5'-m:3'-u] terminus by the LCD is also strengthened by a group of positively charged residues that project side chains into the major groove of the duplex terminus (Fig. 2c).

The MTase domain of HEN1 adopts a core α/β Rossmann structure, in which the cofactor product AdoHcy is bound as in classical S-adenosyl-L-methionine (AdoMet)-dependent MTases²⁴ (Fig. 3b). The ribose ring of AdoHcy directly stacks over the 5'-terminal nucleotide U1_m, and the 5' phosphate of the u strand is hydrogen bonded to the side chain of S747 (Fig. 3b). Three conserved positively charged residues—K749, R753 and K756—interact with the major groove of the [5'-u:3'-m] terminus (Fig. 3b), enhancing the 5' phosphate interaction. These three positively charged residues and S747 are only conserved in plant HEN1 proteins (Supplementary Fig. 7), indicating that recognition of the 5'-phosphate by the MTase domain is not applicable to animal HEN1 homologues. The backbone phosphate connecting the 2-nucleotide 3' overhang and the RNA duplex segment is anchored by a loop formed by six residues (F692 to F697) within motif X of the MTase domain (Fig. 3a). In particular, two non-bridged phosphodiester oxygens form hydrogen bonds with main-chain amines of F692 and L697, respectively (Fig. 3a). The loop structure is further stabilized by hydrophobic stacking interactions between the side chains of F693 and L697 and by the hydrogen bond formed by the side-chain amide of the invariant residue Q700 and the carboxyl oxygen between tandem prolines P695 and P696 in the loop

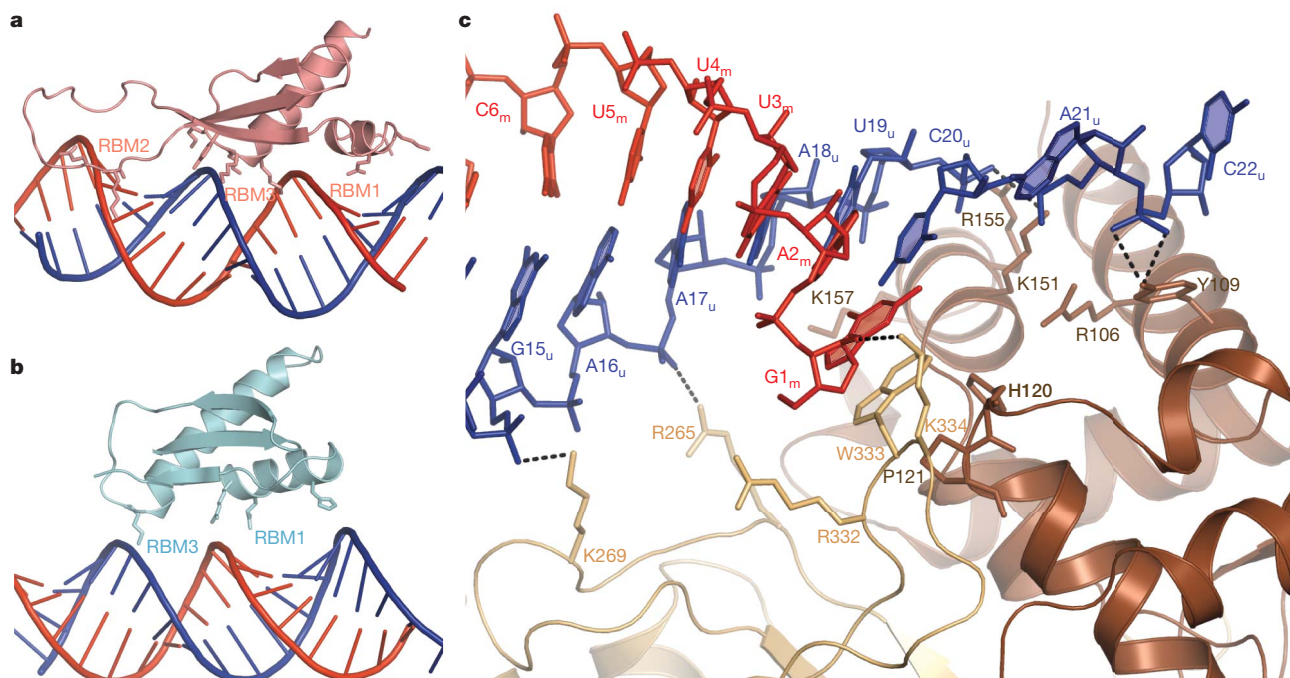


Figure 2 | Small RNA substrate recognition by dsRBDs and LCD. **a**, The duplex region of the small RNA substrate is bound by three RNA binding motifs in dsRBD1. **b**, The duplex region of the small RNA substrate is bound by two RNA binding motifs (RBM1 and RBM3) in dsRBD2. **c**, The LCD

binds to the [5'-m:3'-u] terminus of the small RNA substrate. The 2-nucleotide 3' overhang of the u strand is recognized by the La motif. The base of the 5'-terminal nucleotide G_{1m} is end-capped by W333 in the C-terminal LCD.

(Fig. 3a). Most residues in this loop are invariant in the MTase domains of both plant HEN1 and animal homologues (Supplementary Fig. 7), indicating that this specific interaction by the conserved loop in motif X may also be applicable to animal HEN1 homologues.

The penultimate nucleotide A_{21m} of the 2-nucleotide 3' overhang is flipped out from the duplex and the base of A_{21m} is stacked on the side chains of the conserved residues R856 and L835 (Fig. 3a). The 3'-end nucleotide G_{22m} is flipped back and its base is stacked over the terminal base pair of the duplex (Fig. 3b, c). There are no intermolecular hydrogen bonds between two bases of the 2-nucleotide 3' overhang and the MTase domain, which is consistent with the non-sequence-specific methyltransferase activity of HEN1. The backbone phosphate of the 2-nucleotide 3' overhang is secured by two invariant, positively charged residues R701 and R856 (Fig. 3). Mutation of either R701 or R856 to alanine attenuates the methyltransferase activity of HEN1 (Supplementary Fig. 5b, c), indicating that these two residues are important for the efficiency of HEN1 activity but are not essential.

The ribose ring of G_{22m} is located in the centre of the active site of the MTase domain, where both the 2' and 3' hydroxyls of G_{22m} and the side chains of four invariant residues (E796, E799, H800 and H860) are coordinated to a metal ion, Mg²⁺ (Fig. 3c, d and Supplementary Fig. 8). The highly organized Mg²⁺-mediated coordination precisely presents the 2' hydroxyl of the 3'-terminal nucleotide towards the S_δ atom of AdoHcy (Fig. 3d), indicating that the 2'-O-methylation by HEN1 may be Mg²⁺-dependent (Supplementary Fig. 9). Treatment with increasing concentrations of EDTA that chelates Mg²⁺ in the reaction eventually eliminates HEN1 activity (Supplementary Fig. 5d, f), suggesting that HEN1 is indeed a Mg²⁺-dependent small RNA methyltransferase. Mutation of any one or two coordinated residues to alanines completely abolished HEN1 activity (Supplementary Fig. 5b, c).

Previous biochemical studies^{3,11} defined the features of a small RNA substrate that are strictly required for HEN1 activity: a length of 19–25 nucleotides, a duplex with 2-nucleotide 3' overhang, and

free 2' and 3' hydroxyls on the 3' terminal nucleotide. The structure of the HEN1–small RNA complex revealed that multiple domains in HEN1 cooperate to bind small RNA substrates, which precisely illuminates the RNA substrate specificity of HEN1 (Fig. 4). The RNA substrate may be initially targeted by the N-terminal domain dsRBD1 in HEN1, which allows HEN1 to only act on double-stranded RNAs³. The recognition of an RNA duplex by a classical dsRBD spans about 16 bp²⁵. Thus, the small RNA duplexes produced in plant, approximately 21–24 nucleotides long, are well targeted in the initial recognition. The end-capping interaction by LCD is synergized by dsRBD2, which, together with dsRBD1, forms a strong grip on the duplex region of the small RNA substrate, and these interactions help position the other duplex terminus towards the MTase domain. The recognition of the 2-nucleotide 3' overhang by the MTase domain and the coordination of both the 2' and 3' hydroxyls of the 3'-terminal nucleotide to Mg²⁺ restrict the MTase domain within a limited range where it can efficiently methylate the 2'-hydroxyl on the 3'-end nucleotide. Overall, the preferred length of the small RNA substrates recognized by HEN1 is determined by the distance between the active site of the MTase domain and the 5'-end-capping site in the LCD (Fig. 4).

The mode by which HEN1 measures the length of the small RNA substrate is similar to that of the RNase Dicer, a molecular ruler cleaving the dsRNA substrate at a specified distance from the duplex terminus recognized by the PAZ domain²⁶, although the 3'-end recognition by HEN1 is different compared with that by the PAZ domain^{27,28}. Animal HEN1 homologues only act on single-stranded small RNAs^{4,5,8}, and their small RNA methyltransferase activities are stimulated through interaction with Argonaute proteins⁴ (Y. Kirino, personal communication). Thus, it is possible that animal HEN1 homologues adopt an alternative mode to recognize small RNA substrates (Supplementary Fig. 10), but the mechanism of the Mg²⁺-dependent 2'-O-methylation by the MTase domain is expected to be conserved.

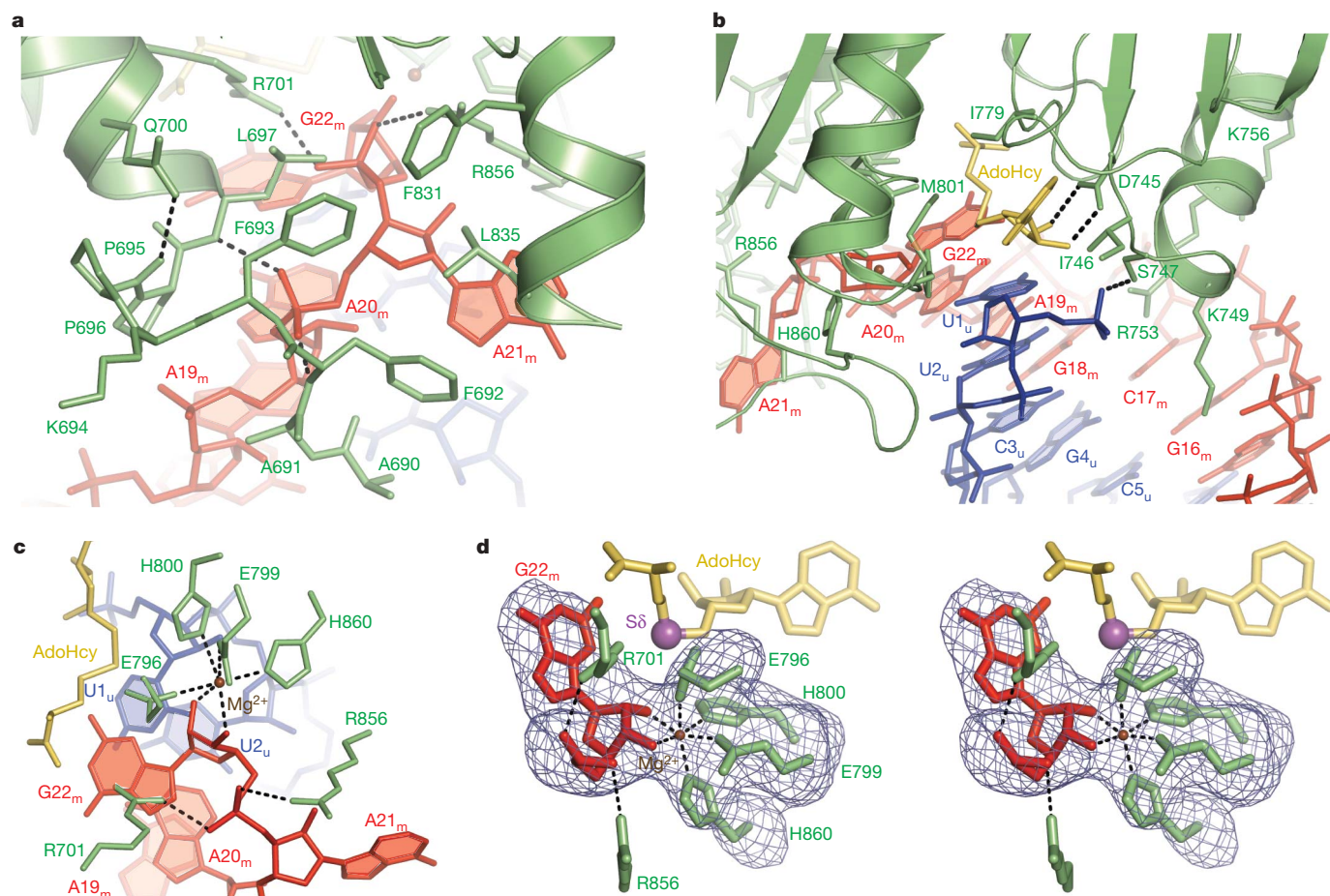


Figure 3 | Small RNA substrate recognition by the MTase domain. **a**, The phosphate connecting the 2-nucleotide 3'-overhang of the m strand with the duplex region is specifically recognized by a conserved loop (F692–L697). The penultimate nucleotide A21_m is flipped out and its base is stacked on the side chains of L835 and R856. The phosphate of the 2-nucleotide overhang is hydrogen bonded by R701 and R856. **b**, The base of the 3'-terminal

nucleotide of the m strand G22_m is stacked on the terminal base pair formed by A20_m and U1_u and the 5'-phosphate of the u strand is recognized by S747. **c**, Both 2' and 3' hydroxyls of the 3'-terminal nucleotide G22_m are coordinated to Mg²⁺ along with four invariant residues, E796, E799, H800 and H860. **d**, A stereo view of the Mg²⁺ coordination covered with $F_o - F_c$ electron density omit map contoured at 3.0 σ .

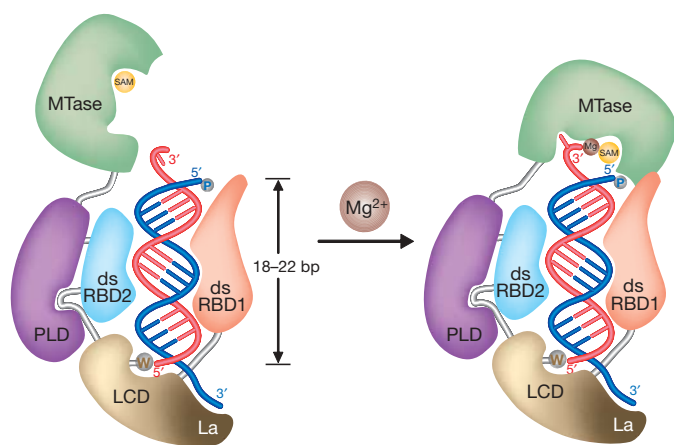


Figure 4 | Proposed model for the specific recognition of small RNA substrates by HEN1 and the Mg²⁺-dependent 2'-O-methyltransferase mechanism. A small RNA substrate is targeted by multiple RNA binding domains in HEN1. The duplex region is gripped by dsRBD1 and dsRBD2, and one terminus is projected towards the MTase domain that is located within a range of 18–22 bp from another terminus end-capped by a tryptophan residue in LCD. Consequently, the MTase domain preferably recognizes the 2-nucleotide 3' overhang on the small RNA substrate of 20–24 nucleotides in length and methylates the 2'-hydroxyl of the 3'-terminal nucleotide in a Mg²⁺-dependent manner.

METHODS SUMMARY

The cDNA of the full-length *Arabidopsis* HEN1 was cloned into the vector pET28 to result in an N-terminal 6 \times His tag and expressed in *Escherichia coli* BL21-Gold(DE3). The protein was purified by affinity, ion exchange and gel filtration chromatography and concentrated. HEN1 mutants were obtained with the QuickChange site-directed mutagenesis kit (Stratagene) or a PCR-based method, and verified by sequencing. RNA oligonucleotides used in the crystallization and assays were ordered from Dharmacon or Integrated DNA Technologies and purified by PAGE or HPLC. Small RNA duplexes were annealed before use. Crystals of HEN1 in complex with the small RNA duplex and AdoHcy were obtained by vapour diffusion with the reservoir solution of 15% PEG3350, 0.2 M sodium chloride, 0.01 M sodium bromide and 0.1 M phosphate-citrate, pH 4.8. The 3.1 Å native data and the 3.4 Å MAD data were collected at beamlines 19BM and 23ID of Argonne National Laboratory, respectively (Supplementary Table 1). The final model was refined on 3.1 Å native data to R_{free} 28.8% and R_{factor} 26.0% with good stereochemistry. Figures were prepared with Pymol (<http://www.pymol.org>).

The *in vitro* small RNA methyltransferase assay was performed as previously described²⁹ with minor modifications. Briefly, 100 μ l methyltransferase reactions were set up for annealed small RNA substrates and HEN1 mutants and monitored by incorporation of the [¹⁴C]methyl group. To assay the Mg²⁺-dependent methyltransferase activity, Mg²⁺ was omitted in all annealing buffer and reactions except as indicated. Different amounts of Mg²⁺ and EDTA were added into individual reactions containing the HEN1 protein and were incubated at room temperature for 15 min before adding small RNA substrates and [¹⁴C]-labelled AdoMet. The *in vitro* RNA–protein crosslinking assay was carried out using iodo-uridine-labelled small RNA substrates as described³⁰.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 16 April; accepted 17 August 2009.

- Ghildiyal, M. & Zamore, P. D. Small silencing RNAs: an expanding universe. *Nature Rev. Genet.* **10**, 94–108 (2009).
- Farazi, T. A., Juranek, S. A. & Tuschl, T. The growing catalog of small RNAs and their association with distinct Argonaute/Piwi family members. *Development* **135**, 1201–1214 (2008).
- Yu, B. *et al.* Methylation as a crucial step in plant microRNA biogenesis. *Science* **307**, 932–935 (2005).
- Horwich, M. D. *et al.* The *Drosophila* RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. *Curr. Biol.* **17**, 1265–1272 (2007).
- Saito, K. *et al.* Pimet, the *Drosophila* homolog of HEN1, mediates 2'-O-methylation of Piwi-interacting RNAs at their 3' ends. *Genes Dev.* **21**, 1603–1608 (2007).
- Kirino, Y. & Mourelatos, Z. Mouse Piwi-interacting RNAs are 2'-O-methylated at their 3' termini. *Nature Struct. Mol. Biol.* **14**, 347–348 (2007).
- Kirino, Y. & Mourelatos, Z. The mouse homolog of HEN1 is a potential methylase for Piwi-interacting RNAs. *RNA* **13**, 1397–1401 (2007).
- Kurth, H. M. & Mochizuki, K. 2'-O-methylation stabilizes Piwi-associated small RNAs and ensures DNA elimination in *Tetrahymena*. *RNA* **15**, 675–685 (2009).
- Chen, X., Liu, J., Cheng, Y. & Jia, D. HEN1 functions pleiotropically in *Arabidopsis* development and acts in C function in the flower. *Development* **129**, 1085–1094 (2002).
- Park, W., Li, J., Song, R., Messing, J. & Chen, X. CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in *Arabidopsis thaliana*. *Curr. Biol.* **12**, 1484–1495 (2002).
- Yang, Z., Ebright, Y. W., Yu, B. & Chen, X. HEN1 recognizes 21–24 nt small RNA duplexes and deposits a methyl group onto the 2' OH of the 3' terminal nucleotide. *Nucleic Acids Res.* **34**, 667–675 (2006).
- Li, J., Yang, Z., Yu, B., Liu, J. & Chen, X. Methylation protects miRNAs and siRNAs from a 3'-end uridylation activity in *Arabidopsis*. *Curr. Biol.* **15**, 1501–1507 (2005).
- Ramachandran, V. & Chen, X. Degradation of microRNAs by a family of exoribonucleases in *Arabidopsis*. *Science* **321**, 1490–1492 (2008).
- Chen, X. A marked end. *Nature Struct. Mol. Biol.* **14**, 259–260 (2007).
- Tkaczuk, K., Obarska, A. & Bujnicki, J. Molecular phylogenetics and comparative modeling of HEN1, a methyltransferase involved in plant microRNA biogenesis. *BMC Evol. Biol.* **6**, 6 (2006).
- Kang, C. B., Dhe-Paganon, S. & Yoon, H. S. FKBP family proteins: immunophilins with versatile biological functions. *Neurosignals* **16**, 318–325 (2008).
- Tian, B., Bevilacqua, P. C., Diegelman-Parente, A. & Mathews, M. B. The double-stranded-RNA-binding motif: interference and much more. *Nature Rev. Mol. Cell Biol.* **5**, 1013–1023 (2004).
- Curry, S. & Conte, M. R. A terminal affair: 3'-end recognition by the human La protein. *Trends Biochem. Sci.* **31**, 303–305 (2006).
- Maraia, R. J. & Bayfield, M. A. The La protein-RNA complex surfaces. *Mol. Cell* **21**, 149–152 (2006).
- Teplova, M. *et al.* Structural basis for recognition and sequestration of UUUOH 3' termini of nascent RNA polymerase III transcripts by La, a rheumatic disease autoantigen. *Mol. Cell* **21**, 75–85 (2006).
- Vargason, J. M., Szittyá, G., Burgyn, J. & Hall, T. M. T. Size selective recognition of siRNA by an RNA silencing suppressor. *Cell* **115**, 799–811 (2003).
- Ye, K., Malinina, L. & Patel, D. J. Recognition of small interfering RNA by a viral suppressor of RNA silencing. *Nature* **426**, 874–878 (2003).
- Yu, B., Chapman, E. J., Yang, Z., Carrington, J. C. & Chen, X. Transgenically expressed viral RNA silencing suppressors interfere with microRNA methylation in *Arabidopsis*. *FEBS Lett.* **580**, 3117–3120 (2006).
- Schubert, H. L., Blumenthal, R. M. & Cheng, X. Many paths to methyltransferase: a chronicle of convergence. *Trends Biochem. Sci.* **28**, 329–335 (2003).
- Ryter, J. M. & Schultz, S. C. Molecular basis of double-stranded RNA-protein interactions: structure of a dsRNA-binding domain complexed with dsRNA. *EMBO J.* **17**, 7505–7513 (1998).
- MacRae, I. J. *et al.* Structural basis for double-stranded RNA processing by Dicer. *Science* **311**, 195–198 (2006).
- Lingel, A., Simon, B., Izaurralde, E. & Sattler, M. Nucleic acid 3'-end recognition by the Argonaute2 PAZ domain. *Nature Struct. Mol. Biol.* **11**, 576–577 (2004).
- Ma, J.-B., Ye, K. & Patel, D. J. Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. *Nature* **429**, 318–322 (2004).
- Yang, Z. *et al.* Approaches for studying microRNA and small interfering RNA methylation *in vitro* and *in vivo*. *Methods Enzymol.* **427**, 139–154 (2007).
- Stump, W. T. & Hall, K. B. Crosslinking of an iodo-uridine-RNA hairpin to a single site on the human U1A N-terminal RNA binding domain. *RNA* **1**, 55–63 (1995).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank K. Sergiy for assistance with the in-house X-ray generator operation, and the staff at Advanced Photon Source beamlines 19ID and 23ID, Argonne National Laboratory, for help with data collection. We thank T. Townes and H. Wang for critical reading of the manuscript. This work was supported by a start-up fund (to J.-B.M.) and partly by a grant from the V Foundation for Cancer Research (to J.-B.M.) and a grant from the National Science Foundation (MCB-0718029 to X.C.). D.G.V. is supported by grants from National Institutes of Health (R01 GM074252 and R01 GM074840).

Author Contributions Y.H. expressed and purified proteins, grew crystals, solved structure and wrote the manuscript. J.-B.M. collected data, solved structure, performed crosslinking assays and wrote the manuscript. L.J. performed small RNA methyltransferase assays and wrote the manuscript. Q.H. and D.G.V. were involved in data processing and refinement. X.C. analysed data and wrote the manuscript. The overall project management and manuscript preparation were by Y.H., X.C. and J.-B.M.

Author Information Atomic coordinates and structural factors for the reported crystal structure have been deposited in the Protein Data Bank under access code 3HTX. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.-B.M. (jma@biochemistry.uab.edu).

METHODS

Protein expression and purification. DNA fragments corresponding to full-length HEN1 were amplified from the cDNA and inserted into the pET28a vector (Novagen) under NcoI and XhoI sites to result in an N-terminal His tag (MGHHHHHH). A double point mutation L604P/K640R was introduced into HEN1 during PCR amplification and L604P was reversed by site-directed mutagenesis to result in the single mutant K640R. Because no differences in methyltransferase activity were observed among the double mutant L604P/K640R, the single mutant K640R and wild-type HEN1 (data not shown), if not specifically indicated, the double mutant L604P/K640R and the single mutant K640R were treated as wild-type HEN1 for protein purification and crystallization in this study. Mutants W333A, E799A/H800A, H860A, R701A and R856A were generated by a PCR-based overlap extension method. E796A, H860Q, Y109A, H800Q as well as the correction of L604P were generated using the QuikChange Lightning site-directed mutagenesis kit (Stratagene). The N-terminal deletion mutant Δ N89 (90–942) was generated by PCR amplification and subcloning. Primers for cloning and mutagenesis are listed in Supplementary Table 2. The presence of the mutations was confirmed by sequencing. The recombinant HEN1 proteins were expressed in *E. coli* BL21-Gold(DE3) (Stratagene). After induction with 0.2 mM IPTG, the cells were allowed to grow at 17 °C for 20 h. Collected cells were lysed by a C-3 cell disruptor (Avestin) at 4 °C. Proteins were purified by affinity His-Trap column, ion-exchange Q column, heparin column and size-exclusive column Superdex 200 (GE Healthcare). Further chromatography on a Mono-Q column (GE Healthcare) was required to obtain high-quality proteins for crystallization. Purified proteins were concentrated to 15 mg ml⁻¹ in a buffer containing 10 mM HEPES (pH 7.5), 50 mM KCl and 2 mM dithiothreitol (DTT), and flash frozen in liquid nitrogen before storing at -80 °C. SeMet-labelled proteins were produced by inhibiting endogenous methionine biosynthesis³¹ in M9 minimal media supplemented with specific amino acids as well as SeMet, and purified as for the native protein. All mutants were expressed and purified as described above.

RNA preparations. Sequences of RNAs used in this study are listed in Supplementary Table 3. All RNA oligonucleotides were synthesized from Dharmacon or Integrated DNA Technologies and further purified with PAGE or HPLC. The concentrations of the RNAs were measured by ultraviolet spectrometry at 260 nm; RNA duplexes used for crystallization and the cross-linking assay were first annealed in a buffer of 30 mM HEPES-K, pH 7.5, 100 mM potassium acetate, 2 mM magnesium acetate. RNA duplexes used for small RNA methyltransferase assay were annealed in a buffer of 50 mM Tris-HCl, pH 7.6 and 100 mM KCl. Annealing was performed by heating the mixture for 5 min at 95 °C and slowly cooling it to 37 °C followed by incubation for 2 h at 37 °C and 1 h at 24 °C in a thermal cycler. Annealing efficiency was examined by running the anneal products on a 15% polyacrylamide native gel.

Crystallization and data collection. Both miR173/miR173* and miR173/miR173*cm RNA duplexes were used to co-crystallize with HEN1 in the presence of AdoHcy. Only miR173/miR173*cm gave out crystals with enough quality for data collection. The ternary complex used for crystallization was prepared by adding 20-fold excess of AdoHcy to HEN1 protein and incubating on ice for 0.5 h followed by the addition of twofold excess of RNA. The final concentration of HEN1 in the complex is about 5 mg ml⁻¹. The initial screening was carried out with commercial crystallization kits using Phoenix crystallization robot (Art Robbins Instruments) and detected using Rock Imager 2 and Rock Maker automated imaging system (Formulatrix). The preliminary hit was obtained at condition number 36 of Wizard II Screen (Emerald Biosystems), 0.1 M phosphate-citrate, pH 4.2, 10% PEG 3000 and 0.2 M NaCl. The crystals were optimized using the hanging-drop vapour diffusion method at 20 °C and Additive Screen (Hampton Research) was used during the optimization of initial condition. Addition of NaBr (Additive Screen, No. 29) markedly improved the quality of the crystals. Finally, the crystals were grown in the solution containing 0.1 M phosphate-citrate, pH 4.8, 15% PEG 3350 and 0.2 M NaCl and 0.01 M NaBr. SeMet-labelled crystals were obtained under the same condition as for the native crystals. Crystals were transferred into cryoprotectant solution with 20% glycerol and then flash-frozen in liquid nitrogen. Diffraction tests of collected crystals were performed at 100 K using a Rigaku X-ray generator equipped with R-Axis IV++ detectors. A multi-wavelength anomalous dispersion (MAD) data set to 3.4 Å was collected on a SeMet-labelled crystal at the Argonne National Laboratory beamline 23ID-B. A native data set to 3.1 Å was collected at beamline 19BM, the Structural Biology Center at the Applied Photon Source. The diffraction data were processed and scaled with the HKL2000 package³². The data collection and processing statistics are summarized in Supplementary Table 1.

Structure determination and refinement. Phase was determined by the multiple-wavelength anomalous dispersion method³³ using 3.4 Å MAD data by

PHENIX package³⁴. Out of a total of 26 sites, 18 selenium atoms were located using the program HYSS in PHENIX³⁴. Heavy atom refinement and MAD phasing were carried out using programs SOLVE and RESOLVE in PHENIX³⁴, and the figure of merit after phasing improvement by program RESOLVE was increased to 0.74 from initial 0.36. An initial model of HEN1 was manually built with the programs O³⁵ and Coot³⁶ using the locations of SeMet positions as guides. The model of the small RNA duplex was built based on the position of the 5'-phosphate that only exists in the miR173 strand. The initial model of the complex was refined through alternating cycles using the program phenix.refine in PHENIX. Non-crystallographic symmetry was used to restrain the core of two domains in the asymmetric unit while more variation was allowed in the loop regions. The final model was refined to the native data in the resolution range 20–3.1 Å using CNS version 1.2 (ref. 37) until the R/R_{free} were 26.0/28.8 with good stereochemistry. Ramachandran analysis showed that 87.2% of residues are in most favoured regions, 12.8% of residues are in additional allowed regions, and no residues in the generally allowed or disallowed regions. The final model contains two double mutant L604P/K660R HEN1 molecules including residues 1–6, 213–215, 290–301, 411–454, 501–534, 542–551, 572–599, 839–850, 912–916 and 934–942 of chain A and residues 1–6, 292–303, 410–452, 501–519, 529–534, 540–546, 574–598, 840–851, 910–917 and 934–942 of chain D (Supplementary Fig. 1), two miR173/miR173*cm duplexes, two AdoHcy, two Mg²⁺, and 37 waters (Supplementary Table 1). Residue 535–541 of chain A and residue 535–539 of chain D are not certain owing to the poor electron density in the middle of a long disordered region. The refined structure was validated using PROCHECK³⁸. Structural figures were prepared with PyMol (<http://www.pymol.org>).

Small RNA methyltransferase assay. The *in vitro* small RNA methyltransferase assay monitored by the incorporation of the [¹⁴C]-methyl group was performed as previously described²⁹ with minor modifications. A 100-μl methyltransferase reaction was set up for the annealed small RNA substrate miR173/miR173* and HEN1 mutants. The reaction mixture contained 50 mM Tris-HCl (pH 8.0), 100 mM KCl, 5 mM MgCl₂, 0.1 mM EDTA, 2 mM DTT, 5% glycerol, 2 μl RNasin (40 U μl⁻¹; Promega), 0.5 μCi S-adenosyl-L-[methyl-¹⁴C] methionine (58.0 mCi mmol⁻¹; Amersham Pharmacia Biosciences), 5 μg purified protein, and 1 nmol RNA substrate. After incubation at 37 °C for 2 h, the reaction was stopped by adding 100-μl 2× proteinase K solution (100 mM Tris-HCl, pH 8.0, 10 mM EDTA, 150 mM NaCl, 2% SDS, and 0.4 mg ml⁻¹ proteinase K) followed by incubation at 65 °C for 15 min. The reaction was then extracted with phenol/chloroform. To precipitate the small RNAs, 1 ml glycogen (5 mg ml⁻¹), 0.1 vol of 3 M NaOAc (pH 5.2), and 2.5 vol of ice-cold 100% ethanol were added to the reaction. The mixture was stored at -80 °C for 2 h and centrifuged at 4 °C for 30 min. The pellet was washed with 100 ml 70% cold ethanol. The RNAs in the pellet were dissolved with 1× RNA loading buffer, heated at 95 °C for 5 min, immediately put on ice, and loaded on to a 15% denaturing polyacrylamide gel with 7 M urea. After electrophoresis, the gel was treated with an autoradiography enhancer (Enhance from Perkin Elmer) following the manufacturer's instructions and exposed to X-ray film at -80 °C. To assay the Mg²⁺-dependent methyltransferase activity, Mg²⁺ was omitted in all annealing buffer and reactions except as indicated. Different amounts of Mg²⁺ and EDTA were added into individual reactions containing the HEN1 protein and were incubated at room temperature for 15 min before adding small RNA substrates and [¹⁴C]-labelled AdoMet.

The *in vitro* RNA-protein photochemical crosslinking assay. Small RNA miR173 with a 5-iodouracil (5IU) substitute at U1 or U21 (mutated from A21) (Supplementary Table 3) was 5'-end labelled using γ-³²P-ATP (NEN), and annealed with miR173*cm or used as a single-stranded substrate. The photochemical crosslinking assay was performed as described³⁰. Typically, 20-μl reactions containing 0.2 μM RNA and 2 μM HEN1 or mutants were placed in a 1.5-ml microtube and incubated for 20 min on ice. Exposure to the ultraviolet light source (Spectrolite, λ_{max} = 312, 330 μW cm⁻²) was at a distance of ~2.5 cm, filtered through a polystyrene Petri dish for 10 min. Crosslinking products added with 2× loading buffer were separated on 12% SDS-polyacrylamide gels, which were exposed to Storage Phosphor Screen (GE Healthcare) and visualized using a Storm PhosphorImager (GE Healthcare).

31. Doublé, S. & Carter, C. W. Jr. Preparation of selenomethionyl proteins for phase determination. *Methods Enzymol.* **276**, 523–530 (1997).
32. Otwinowski, Z., Minor, W. & Carter, C. W. Jr. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
33. Hendrickson, W. A. Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* **254**, 51–58 (1991).
34. Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58**, 1948–1954 (2002).

35. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
36. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
37. Brunger, A. T. Version 1.2 of the Crystallography and NMR system. *Nature Protocols* **2**, 2728–2733 (2007).
38. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291 (1993).

ERRATUM

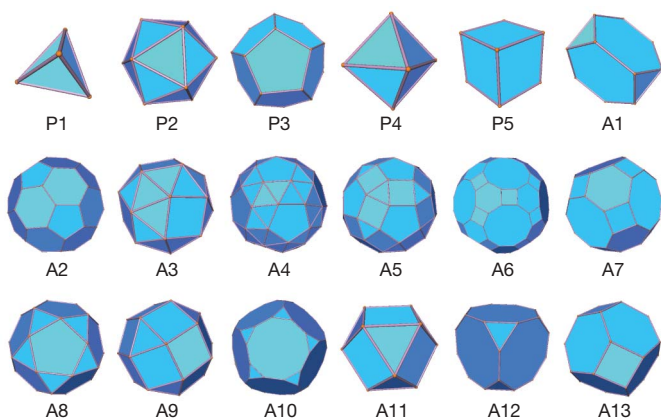
doi:10.1038/nature08492

Dense packings of the Platonic and Archimedean solids

S. Torquato & Y. Jiao

Nature 460, 876–879 (2009)

In Figure 1 of this letter, in the top row 'A1' was incorrectly listed as 'P6'. The correct figure is shown below.



ERRATUM

doi:10.1038/nature08493

Stable single-unit-cell nanosheets of zeolite MFI as active and long-lived catalysts

Minkee Choi, Kyungsu Na, Jeongnam Kim, Yasuhiro Sakamoto,
Osamu Terasaki & Ryong Ryoo

Nature **461**, 246–249 (2009)

In this Letter, the affiliations for Osamu Terasaki were incorrect. This author is associated with affiliations 4 and 5 from this Letter, the Graduate School of EEWS (WCU), KAIST, and Structural Chemistry, Arrhenius Laboratory, Stockholm University, and not with Nanoscience and Nanotechnology Research Center, Osaka Prefecture University.

CORRIGENDUM

doi:10.1038/nature08523

**Genotypic sex determination enabled
adaptive radiations of extinct marine reptiles**

Chris L. Organ, Daniel E. Janes, Andrew Meade & Mark Pagel

Nature 461, 389–392 (2009)

In this Letter, '*Dolichorhynchops osborni*' was incorrectly listed as '*Dolichorhynchops osburni*' at two occasions in the text.

Life in a monastic lab

A vocational career.

Joost Uitdehaag

The bell rang for evensong as Jorge attached the power-pack and started his gel. He smiled. He liked it when everything was exactly in time. He left the lab and walked towards the chapel. On the way he met his older friend, Anselm, who hurried along as usual.

"Slow down," Jorge whispered. "What's the use?"

"What's the use of being slow?"

"Slow is about taking aim."

"Where did you get that from?"

"A penitence session."

"Don't mention those."

"You mean they are counterintuitive to a fearful old individualist. Really, you should join. Maybe even tonight?"

Anselm just smiled. They stopped talking as they entered the chapel. It had a pleasing retro ambience — its design influenced by Le Corbusier's famous Chapel of Notre Dame — amid the lab complex of the Benedictine Order for Oncology, set in a remote valley in the Ardennes.

For Jorge, his lab was one of the good things the great crisis had brought: a total reshuffling of drug research, an injection of idealism in a world of self-interest. That the injection had come from religion was no surprise for Jorge. Management gurus had been courting religious rules long before the crisis. Live for yourself or for your community, that was the post-crisis choice, and science and religion were both community efforts. Scientific monasticism was a new synthesis, the ultimate way of serving society.

All the scientists had gathered in the chapel, and they started a medieval hymn. Singing together was supposed to stimulate collaboration and equality, but Jorge was still bad at it. During the hymn he worried about Anselm. His friend had started to complain again about giving up the 'self' side of science. He was a former academic and had this all-pervading desire to compete and establish his name, but within the Order that would get him into trouble. They gave you a permanent contract and a budget so there was no need to worry about grants or tenure, but in return the Order demanded no double work, no egos and no secrecy.

If only Anselm had been a pharma man. Novices from industry generally had less trouble giving up the self-side. But then again, those who had worked through the Barren Years had generally less passion for their jobs than a zebrafish for a barcode.

Jorge wondered why people could not simply decide if they really wanted to live their undergraduate dreams and work on curing disease, or if they wanted something else. Anselm always said he was naive.

"Idealists have a history of getting hurt," he would say.

"Isn't that the whole point," Jorge would reply, "that contributing costs you?"

"You just haven't suffered yet."

Anselm had been damaged by his time in academia; that much Jorge knew. That's why doing penitence tonight would be good for him. It would give him that perfect feeling that all was well and that he



was living a good life. If only Jorge could convince him.

The singing finished and Abbott Fra Paolini spoke about the Barren Years. That was the time when ever larger pharma companies and a society ever more hostile to them together had driven the cost of developing a drug to \$2 billion. And what was considered worse: to the cost of a thousand scientific careers. It had been the scientific equivalent of the Somme offensive.

With a wide movement of his hands, Fra Paolini spoke of the day when seven ex-pharma scientists had taken up vows in a monastery to continue a 'killed' project. It was a golden move. Their vows of poverty (no patenting, no bonuses), chastity (do nothing that satisfies only yourself) and obedience (listen to what patients want) were the right guarantees for patient organizations and health insurers to pour money into monastic research labs. In the past year,

these labs had developed and published the majority of new therapies (generics companies usually took up marketing them).

After the ceremony, Jorge waited for Anselm.

"Why do you fear a penitence session? You know they did this all the time in the old days: remember Borel and cyclosporine? It's part of our tradition. It is why the public likes us."

"I don't fear it. I just don't think it's rational. It's hysterics."

"I'm not hysterical."

"But you're not joining tonight are you?"

Jorge did not answer. Anselm stopped walking and gave him an angry look.

"You are! That would be what, the second time in a month? You're wasting yourself."

"The supervisory committee allowed me."

"Sure they do. Bunch of vampires, they are."

"It has nothing to do with them and all with me," Jorge hated being berated.

"I won't allow you," Anselm said.

"What do you want to do? Swap places?"

"If that's what it takes."

Jorge was amazed. Was getting Anselm to do penitence really this simple? Was he really going to give up his principle for a worry about a friend? Anselm never ceased to surprise him.

"All right," he said.

In the monastery, most clinical trials were carried out in the infirmary, on a veranda filled with the evening's sunlight. Jorge was sitting at Anselm's bed.

"You are getting chimidininib," Jorge said, "the first inhibitor of the Chung-Mi variant isomerase. Have you seen the preclinical data?"

"Yes. They're ok."

Jorge rolled up Anselm's sleeve as a nurse prepared the drip.

"You want some blood for western blotting tomorrow?"

Anselm nodded. "Don't worry," he said as the compound started to enter his body.

But Jorge felt guilty. That night he did the only sensible thing: he lit a candle for his friend.

Joost Uitdehaag lives in the Netherlands and works for a drug-discovery company. His writing includes literature on drug discovery and fantasy stories for Dutch magazines.

Join the discussion of Futures in Nature at <http://tinyurl.com/kkh3kt>

Q&A

Robert Margolskee, an expert in the molecular mechanisms of taste, has recently accepted a faculty position at the Monell Chemical Senses Center in Philadelphia, Pennsylvania.



How did your interest in the molecular biology of taste evolve?

In high school, I was intrigued by molecular-biology techniques and their potential in conducting neuroscience studies of brain function. In college and medical school, my professors questioned my interest in a field that didn't yet exist. But as a postdoc I saw colleagues use molecular techniques to study vision and smell. I realized taste was a complete 'black box': we had no idea how taste cells worked at the molecular level. So I saw it as an opportunity to identify and clone taste receptors.

What was your first 'aha' moment?

In late 1991, we found a new protein in taste cells that was closely related to the protein transducin, which transmits visual signals to the brain. At first we took this new insight into the taste system to mean that if transducin-like proteins were in taste cells, taste might be closely related to other sensory systems

such as vision. We found that although the taste version of transducin (ultimately called gustducin) was structurally similar to transducin, only the signalling outputs of the two were similar.

What do you consider to be your greatest scientific achievement?

We molecularly characterized gustducin's involvement in sweet, bitter and umami (the monosodium glutamate taste). That is a stepping stone to further studies of taste signalling elements, which we have found are expressed elsewhere in the body and contribute to non-taste functions in the stomach and pancreas. At Monell, I plan to circle back to the role of these proteins in health and disease.

Do you get bombarded with questions from the food industry?

No, but I have been in some interesting forum discussions with molecular neuroscientists and chefs. It is interesting to compare

notes, and at some point I'm sure we'll get to the level of understanding how taste works to apply it to the creation of a meal or dessert.

What do you value most about the scientific process?

There is a purity and a clarity in discovery and publication that is closely related to nature and truth.

What is the key to navigating a successful scientific career?

I wish I knew. I guess it is a matter of balance. You have to balance everything in your life — personal and professional, bench work and supervising others, what appeals to you and what will get funded. To be effective and successful, you must find a way to follow your heart and anticipate what the journal editors will say. But in the end, I think you can approach your career in different ways and end up coming to the same point. ■

Interview by Virginia Gewin

IN BRIEF

Endangered papers

Conservation scientists take up to three times longer to publish their work than other biologists, according to a new study that warns this could affect time-critical environmental decisions. Ryan O'Donnell and two other PhD students at Utah State University in Logan examined more than 2,000 articles published in 14 life-science journals in 2007 to calculate the delay between last data collection and submission. The median delay for papers on conservation was 696 days, compared with 189 days for evolution and 605 for taxonomy. The authors suggest the hold-up arises because many conservation biologists do governmental work and have other obligations besides publishing.

Bridges to biotechnology

Oregon's engineers and other skilled workers who lost their jobs in the economic downturn have a new alternative, thanks to a recently established biotechnology retraining scheme. The Bioscience Foundations Program, which is jointly funded by a US\$136,000 federal-stimulus grant and the state of Oregon, aims to match people to short internships at Portland-based bioscience companies — particularly medical-device firms — that could lead to permanent positions. Internship applicants will be interviewed by the firms themselves; those selected will be tutored by bioscience experts on process and compliance in the industry, as well as on issues of policy, environment, ethics, and research and development.

Gender imbalance persists

Men continue to make up the majority of doctoral scientists and engineers in the United States, according to the US National Science Foundation's most recent 2006 Survey of Doctorate Recipients, released on 24 September. The report says that men comprise some 68% of America's 561,230 doctoral scientists across all science fields and 90% of the nation's 121,520 doctoral engineers. The last survey, conducted in 2003, found that men represented 70% of doctoral scientists and 91% of doctoral engineers. Of the 10,920 doctoral scientists and engineers across all fields who are neither working nor seeking work, three-quarters are female, the latest report says, up from the two-thirds reported in the 2003 survey.

POSTDOC JOURNAL

Communicating science



Take one scientist. Blend in professional science communicators. Incubate in crisp mountain air. The result: an ability and a desire to discuss science with all sorts of audiences.

Two months ago I'd never made a film, designed a website or written a science news piece. By the end of August I'd had a major part in all three. How? I was fortunate to participate in an intensive two-week science-communication programme at the Banff Centre, a crucible of literary and performance art nestled in the Rocky


Mountains of Alberta, Canada.

Participants receive hands-on training with a broad range of media. Our teachers included accomplished professionals in television, radio, web and print journalism. Their mentorship, combined with audio and film equipment and web-design support, taught me how to talk science successfully via multiple media. Our group, for example, made lively podcasts, which we called 'Bunk Debunk', to define scientific jargon in clear terms.

I attended the programme, in part, to improve on the

campus-radio science show that I host. Now I feel invigorated — able and eager to talk science effectively, not only through radio and podcasts but also via television and print media. My passion for public science communication nearly matches that for my research. Now I'm certain that I want to nurture a career in science communication, whether as a sideline to my research or, perhaps, as my full-time job. ■

Julia Boughner is a postdoc in evolutionary developmental biology at the University of Calgary, Canada.



'Skywalk' in downtown Nagoya.

RISING STAR?

The Japanese city of Nagoya is aiming to turn a history of manufacturing success into fertile ground for science applications. **David Cyranoski** surveys its potential.

The car-maker named after its home city of Toyota has brought international fame and economic vitality to Japan's central coastal region, which has the city of Nagoya at its heart. The area's industrial prowess has made it the country's most productive manufacturing zone for the past three decades. But in terms of international standing in science and technology, Nagoya — Japan's fourth most-populous city — and its university remain under-appreciated and largely unknown.

"Everyone knows Toyota, but people don't know Nagoya or the university," laments Nagoya University's vice-president Takashi Miyata. Founded in 1939 as one of Japan's seven prestigious imperial universities and with roots as a medical-training institution in the 1870s, Nagoya University has nowhere near the international visibility of national universities in Tokyo, Kyoto and Osaka. It stands at 120 in the 2008 *Times Higher*

Education Supplement rankings — the other three are all in the top 50. And yet four of the seven Japanese scientists to win a Nobel prize this century were graduates or staff of the university: chemists Ryoji Noyori and Osamu Shimomura, and theoretical physicists Toshihide Maskawa and Makoto Kobayashi.

Now the university and city and regional governments are planning to transform Nagoya into a research hub. They will focus on the region's strong point: *monozukuri*, which, literally translated, means 'making things'. But the term also implies integrated production and the ability to put things together in a creative fashion, the way that Toyota excels at putting together cars, says Miyata.

Nagoya University crystallographer Yoshikazu Takeda describes *monozukuri* as "manufacturing with craftsmanship". He says it requires practitioners to "feel by their finger a difference in thickness as small as one micrometre".

Scientists and science policy-makers are

now striving to modernize *monozukuri*, using academic strengths in physics and chemistry, as well as a state-of-the-art plasma-research centre and synchrotron facilities. If all goes according to plan, they will bring Nagoya's science to a new level of international visibility and economic fortune.

Monozukuri makes good

Efforts to develop applied science with industry applications are already poised to bear fruit. For example, Nagoya University's Center for Embedded Computing Systems was established in 2006 and is supported through collaborative research projects with several companies, including Toyota. In a project co-developed with nearby AutoNetworks Technologies, the centre has applied for eight patents for technology that would standardize the networks of a car's 80-plus electronic control units. Projects with Toyota include the development of a new multimedia system that would enable, for example, remote diagnosis of car problems.

But the *monozukuri* extends far beyond cars. Central government set up one of its 18 'knowledge clusters' in the region in 2003. Called the Nagoya Nanotechnology Manufacturing Cluster, it includes Nagoya University, Nagoya Institute of Technology, Toyohashi University of Technology and Meijo University, as well as 56 local companies in the quest "to make environmentally friendly, highly functional materials that lead the world". It aims to bring together expertise, foster collaborations and seek out industrial applications for science projects, with a focus on plasma nanotechnology. The cluster already includes the Plasma Centre for Industrial Applications, which started operating last August to help introduce new plasma technologies, especially to small and medium-sized *monozukuri* firms. It was set up with matching funds from the Aichi prefecture, where Nagoya is located.

Plasma, a partially ionized gas known as the fourth state of matter, has valuable properties. For example, its free radicals react with a substrate surface, enabling the etching of patterns for integrated circuits, even at room temperature. But finding the best match between a given surface, the desired pattern and depth and a plasma remains difficult.

"Plasma scientists don't really know what's happening," says Keigo Takeda, a researcher on the project from Nagoya University. Generally speaking, the process is mostly trial-and-error. "There are huge amounts of waste," says Takeda.

Takeda and his colleagues, led by Masaru Hori, have invented a device that not only can tell the type and concentration of the radicals in the plasma, but can also guide a self-adjusting feedback mechanism to ensure stable quality during production. "We're aiming for a device that anyone can use, one that can really speed up research and

development,” says Takeda. Hori’s team is now testing new applications, including the processing of materials for parts used in cars or aeroplanes. The plasma can be used to add a film that hardens the material or makes it easier to paint.

Nanotech cluster projects are already spinning off companies and collaborations. Hori’s device is being developed by a spin-off called NU Eco Engineering. Nagoya University’s Osamu Takai is also commercializing his biomimetic self-assembling monolayers — molecules that, under certain conditions, organize themselves into layers based on hydrophilic or hydrophobic properties. The super-hydrophobic films, created with plasma processing, are being manufactured by Nagoya-based spin-off n-Factory. The technology could be used in automobile parts, medical equipment, and DNA and protein chips, as well as in photoresistors in semiconductor manufacturing. The cluster has also given rise to Meijo Nano Carbon, which sells carbon nanotubes.

Funding renewed

In October 2008, the central government made this one of only nine clusters to be renewed for another five years, earning it ¥1 billion (US\$11 million) a year from the central government, to be matched by ¥500 million a year from city or prefectural governments.

The government’s decision was partly influenced by the training of industry-minded scientists and collaborations with foreign and domestic partners. For example, Nagoya University’s Plasma Nanotechnology Research Center has ties with ten foreign research institutions. Infrastructure development also spurred government backing. Largest among these projects is the Central Japan Synchrotron Radiation Research Facility, which has received ¥20 billion in funding from the Aichi prefectural government. Based east

of Nagoya and expected to start operations in 2012, the synchrotron will swallow ¥8 billion of the budget. It joins SPring-8 (the world’s most powerful light source) in Hyogo to the west and Photon Factory northeast in Tohoku, but unlike these, it will have commercial aims. “It will be built to the specifications, in terms of reliability and stability, needed for industrial research and development,” says Yoshikazu Takeda.

The potential users are many — starting with Toyota. The company has used SPring-8 to help design exhaust gas catalysts, hybrid car batteries for its Prius, and fuel cells. “But the rigorous specifications will be a boon for basic researchers who are able to use the facility,” says Takeda. He says it will cover 90% of the research topics possible at SPring-8 and will have some capabilities that SPring-8 does not have. For example, the new synchrotron will have capacity for long-term studies looking at 1,000 samples in a series to ensure reliability of industrial products and reproducibility in basic-science experiments.

Regional policy-makers hope that the infrastructure investment will attract more venture capital and encourage more university spin-offs. Aichi had 74 university-driven business ventures in 2007, placing it in sixth place in Japan. Masaki Kato, of the Aichi New Industry Division’s science and technology section, says they could do better. “There is less capital here than in Tokyo, for example, but because of all the industry there are big opportunities,” he says. Although the recent recession has caused a drop in capital,



Keigo Takeda: cutting waste.



Osamu Takai: commercial goals.

Kato notes that many venture funds have been established, such as Aichi Venture House, and he is encouraged by a growing interest in industry among university researchers.

Bioscience efforts also tap the local expertise in nanotechnology and engineering. Nagoya University’s Innovative Research Center for Preventive Medical Engineering was established in 2006 to bridge the fields of medicine and engineering with a budget of ¥1.7 billion for four years (half from the science ministry and half from local companies). A group led by the centre’s chemist, Yoshinobu Baba, for example, developed a method for separating out cancer stem cells in a blood sample, and other groups in the centre found new markers to predict metabolic syndrome and developed techniques to diagnose milk allergies.

Hoping to commercialize his technology, Baba will start working with Nagoya University Hospital, which is part of a medical network with more than 20,000 beds. He is confident that commercialization will move swiftly. “There are many *monozukuri* companies that want to collaborate with my group,” he says. The centre is now applying for a ¥4.2-billion renewal of its grant for the 2010–16 period.

Nagoya still lacks a significant biotech or pharmaceutical presence, especially since Pfizer cancelled a planned Nagoya-based research arm in 2007 as part of worldwide cutbacks. There is, however, some hope: the unit’s intellectual property was spun off as a biotechnology company named RaQualia Pharma, which has set its sights on an initial public offering in 2011.

But to make good on its international aspirations, the city will need a new and diverse crowd of scientists. Baba and others have foreign postdoctoral students, but attaining more senior positions remains difficult for foreigners, largely because of language barriers, says Yoshikazu Takeda. Nagoya University has set up liaison offices in eight cities around the world to help with recruitment. But even if it continues to struggle to match the scientific reputation of Japan’s better-known hubs in greater Tokyo and the Kyoto region, current efforts to enhance its capacity for *monozukuri* could soon make Nagoya a major player.

David Cyranoski is Nature’s Japan correspondent.



Nagoya University: home to four Nobel laureates and now hoping to raise its profile worldwide.

Gal4 turnover and transcription activation

Arising from: K. Nalley, S. A. Johnston & T. Kodadek *Nature* **442**, 1054–1057 (2006)

Growing evidence supports the notion that proteasome-mediated destruction of transcriptional activators can be intimately coupled to their function^{1,2}. Recently, Nalley *et al.*³ challenged this view by reporting that the prototypical yeast activator Gal4 does not dynamically associate with chromatin, but rather ‘locks in’ to stable promoter complexes that are resistant to competition. Here we present evidence that the assay used to reach this conclusion is unsuitable, and that promoter-bound, active Gal4 is indeed susceptible to competition *in vivo*. Our data challenge the key evidence that Nalley *et al.*³ used to reach their conclusion, and indicate that Gal4 functions *in vivo* within the context of dynamic promoter complexes.

Studies by several groups, including ours^{1,2,4–6}, have reported an intimate connection between the activity of transcriptional activators such as Gal4 and Gcn4 and their destruction by ubiquitin-mediated proteolysis. This intimate connection is difficult to reconcile with the conclusion by Nalley *et al.*³ that proteolytic turnover of Gal4 is not coupled to its function. This conclusion is based on the result of chromatin immunoprecipitation (ChIP) experiments showing that endogenous, active, Gal4 cannot be competed from the *GAL1/10* promoter by induction of a protein with the same DNA-binding specificity (‘competitor’). In this case, the competitor contains the hormone-binding domain of the oestrogen receptor (ER), which allows its DNA-binding activity to be rapidly induced by the treatment of yeast with β -oestradiol. Yeast cultures expressing both the competitor and endogenous Gal4 are treated with β -oestradiol, and ChIP analysis is used to monitor the binding of the two proteins to the *GAL1/10* promoter.

We obtained reagents from the Kodadek laboratory and repeated their experiments. In the course of performing an additional control that was not included in their *Nature* paper, we observed that, in the absence of any competitor, β -oestradiol induced an up to fourfold increase in the levels of Gal4 that associated with the *GAL1/10* promoter (Fig. 1a, blue line). The unexpected ability of β -oestradiol to induce binding of endogenous Gal4 makes the competition assay difficult to interpret, as the compound is simultaneously inducing both the competitor and the species being competed.

To explore this issue further, we repeated the experiment using a different ER ligand, 4-hydroxytamoxifen (4HT). In the absence of competitor, 4HT had little effect on the association of endogenous Gal4 with its cognate promoter (Fig. 1a, red line). Consistent with the different effects of these two ligands on basal association of Gal4 with chromatin, the two compounds gave very different results in the presence of competitor (Fig. 1b). As Nalley *et al.* published³, the addition of β -oestradiol to yeast expressing the competitor protein resulted in little if any reduction in the levels of endogenous Gal4 at the *GAL1/10* promoter, creating the impression that most promoter-bound Gal4 resisted competition (Fig. 1b, blue line). In the presence of 4HT, however, the opposite result was obtained, and $\sim 75\%$ of endogenous Gal4 was competed from the chromatin within 15 min of the ligand addition (Fig. 1b, red line). Notably, the loss of Gal4–chromatin association was accompanied by loading of the competitor onto the *GAL1/10* promoter (Fig. 1c), consistent with the notion that the 4HT-activated competitor can displace endogenous Gal4 from the promoter. Although the competitor protein associated with the *GAL1/10* promoter with apparently slower kinetics than endogenous Gal4 dissociated (compare red line in Fig. 1b with pink line in Fig. 1c), it is worth noting that endogenous Gal4 can bind cooperatively to several sites *in vivo*⁷. There are four Gal4-binding sites in the *GAL1/10* promoter. Thus, a single competitor bound to one of the sites could have the effect of destabilizing multiple Gal4–promoter complexes, leading to efficient displacement of endogenous Gal4 at substoichiometric levels of competitor.

On the basis of our observations, we propose that the recalcitrance of Gal4–promoter complexes originally reported by Nalley *et al.*³ is an artefact of using β -oestradiol to stimulate the competitor. Activating the competitor with 4HT (Fig. 1b), or normalizing the β -oestradiol signal to the important ‘no competitor’ control (Fig. 1d), shows that Gal4 can indeed be rapidly displaced from promoter DNA *in vivo*. Their conclusion that Gal4–promoter complexes lock in and have long half-lives under activating conditions is thus unsustainable.

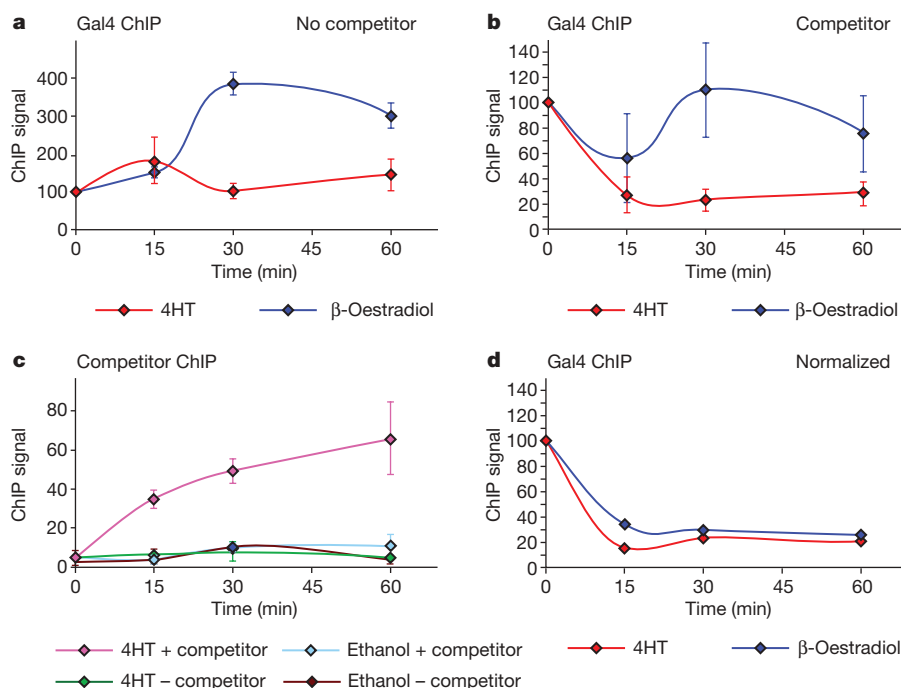


Figure 1 | Activation of a Gal4 competitor with β -oestradiol versus 4HT. **a**, Wild-type yeast were induced with 2% galactose for 60 min and β -oestradiol or 4HT was added. At indicated times, the occupancy of endogenous Gal4 on the *GAL1/10* promoter was determined by ChIP. ChIP signal is normalized to that at time zero. **b**, As in **a**, except that experiment was performed in yeast expressing the Myc-G4-ER-VP16 competitor (supplied by T. Kodadek³). **c**, As in the 4HT experiment in **b**, except that ChIP was used to monitor association of the Myc-G4-ER-VP16 competitor with the *GAL1/10* promoter. The corresponding non-competitor controls are also shown. To calculate the percentage binding in this case, ChIP signals were normalized to those from a Myc-G4-ER-VP16 ChIP (60-min time point) performed in the absence of endogenous Gal4, which corresponds to the total amount of competitor that can bind in this assay. **d**, ChIP signals from β -oestradiol or 4HT experiments in **b** normalized to the relevant ‘no competitor’ control in **a**. Error bars are s.e.m. ($n = 3$).

METHODS

Yeast (BY4741) with or without competitor (Myc-G4-ER-VP16)³ were grown in complete synthetic medium (CSM) (2% raffinose) and Gal4 was induced by transferring yeast to media containing 2% galactose for 1 h. Yeast were then treated with 1 μ M 17- β oestradiol (Sigma) or 100 μ M 4-hydroxytamoxifen (Sigma) for the indicated times. ChIP was performed⁵ using either the Gal4-TA C-10 (anti-GAL4; Santa Cruz) or AB1 (anti-Myc; Calbiochem) antibodies. DNA enrichment was calculated as described⁸ using *ACT1* as the reference locus.

Galen A. Collins^{1,2†}, J. Russell Lipford^{3†}, Raymond J. Deshaies³ & William P. Tansey^{2†}

¹Watson School of Biological Sciences, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA.

²Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA.

e-mail: william.p.tansey@vanderbilt.edu

³Caltech, Division of Biology, 1200 East California Boulevard, Pasadena, California 91125, USA.

[†]Present addresses: Vanderbilt University Medical Centre, 465 21st Avenue South, Nashville, Tennessee 37232, USA (G.A.C. and W.P.T.);

Amgen, 1 Amgen Center Drive, Thousand Oaks, California 91320-1799, USA (J.R.L.).

Received 20 February; accepted 13 August 2009.

1. Lipford, J. R. & Deshaies, R. J. Diverse roles for ubiquitin-dependent proteolysis in transcriptional activation. *Nature Cell Biol.* **5**, 845–850 (2003).
2. Muratani, M. & Tansey, W. P. How the ubiquitin-proteasome system controls transcription. *Nature Rev. Mol. Cell Biol.* **4**, 192–201 (2003).
3. Nalley, K., Johnston, S. A. & Kodadek, T. Proteolytic turnover of the Gal4 transcription factor is not required for function *in vivo*. *Nature* **442**, 1054–1057 (2006).
4. Lipford, J. R., Smith, G. T., Chi, Y. & Deshaies, R. J. A putative stimulatory role for activator turnover in gene expression. *Nature* **438**, 113–116 (2005).
5. Muratani, M., Kung, C., Shokat, K. M. & Tansey, W. P. The F box protein Dsg1/Mdm30 is a transcriptional coactivator that stimulates Gal4 turnover and cotranscriptional mRNA processing. *Cell* **120**, 887–899 (2005).
6. Salghetti, S. E., Muratani, M., Wijnen, H., Fletcher, B. & Tansey, W. P. Functional overlap of sequences that activate transcription and signal ubiquitin-mediated proteolysis. *Proc. Natl Acad. Sci. USA* **97**, 3118–3123 (2000).
7. Giniger, E. & Ptashne, M. Cooperative DNA binding of the yeast transcriptional activator GAL4. *Proc. Natl Acad. Sci. USA* **85**, 382–386 (1988).
8. Ezhkova, E. & Tansey, W. P. Chromatin immunoprecipitation to study protein-DNA interactions in budding yeast. *Methods Mol. Biol.* **313**, 225–244 (2006).

doi:10.1038/nature08406

Nalley et al. reply

Replying to: G. A. Collins, J. R. Lipford, R. J. Deshaies & W. P. Tansey *Nature* **461**, doi:10.1038/nature08406 (2009)

Proteasome-mediated turnover of some^{1,2}, but clearly not all^{3,4}, transcriptional activators is important for their activity. To facilitate the analysis of activator–promoter complex lifetime *in vivo*, a parameter relevant to this issue, we developed a competition chromatin immunoprecipitation (ChIP) assay in which binding of a native transactivator to its cognate promoters is challenged by a ligand-activated competitor protein with the same DNA-binding specificity. We applied this technique to the yeast Gal4 system⁵ and concluded that under non-inducing conditions (raffinose media) Gal4–promoter complexes exchange rapidly, but under inducing conditions (galactose media) the activator–promoter complexes are long-lived. Collins *et al.*⁶ report that, surprisingly, the addition of oestradiol to yeast lacking Myc-G4-ER-VP16 increased the amount of DNA co-immunoprecipitated with native Gal4.

This is a control we had not done, but have subsequently repeated and agree that this is the case (S.A.J. wishes to note that he had requested this control and it erroneously was not done). We thank Collins *et al.*⁶ for pointing out this omission. They go on to show that inducing competitor protein activity with 4-hydroxytamoxifen (4HT) results in a significant loss in the intensity of the ChIP signal owing to native Gal4, but that this ligand does not affect the intensity of these ChIP signals in the absence of competitor. They also show significant association of the competitor protein with the promoter, although with a different time course than Gal4 dissociation. We agree that these data indicate that a large fraction of Gal4–promoter complexes are kinetically labile *in vivo* under these (4HT-containing media) conditions. It is important to note that this odd effect of steroid is not a general problem in the application of this technology to the measurement of other activator–promoter half-lives⁴.

However, our ChIP data tracking association of the competitor protein do not support the conclusion of a rapidly exchanging Gal4–DNA complex *in vivo* in the presence of β -oestradiol rather than 4HT. There is no indication that these data are compromised by unanticipated effects of β -oestradiol. Under inducing conditions,

much lower levels of association of the competitor protein with GAL promoters were observed when Gal4 was present than in *Agal4* cells when β -oestradiol was used to trigger the competition. These data argue for the presence of a stable, functional Gal4–promoter complex in the presence of galactose and under the particular conditions used in our study⁵. It may be that the stability of Gal4–promoter complexes are somehow affected by steroid receptor ligands, which would explain the different results observed by ourselves and Collins *et al.*⁶ for the association of the competitor protein in our respective experiments.

Kip Nalley^{1†}, Stephen Albert Johnston^{1†} & Thomas Kodadek^{1†}

¹Department of Internal Medicine, UT Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, Texas 75390-9185, USA.

Email: kodadek@scripps.edu

[†]Present addresses: Laboratory of Receptor Biology and Gene Expression, National Cancer Institute, Building 41, Room B602, 41 Library Drive MSC 5055, Bethesda, Maryland 20892-5055, USA (K.N.); Center for Innovations in Medicine, Biodesign Institute, Arizona State University, Tempe, Arizona 85287-5001, USA (S.A.J.); Departments of Chemistry & Cancer Biology, Scripps Research Institute, Scripps Florida, 130 Scripps Way, Jupiter, Florida 33458, USA (T.K.).

1. Reid, G. *et al.* Cyclic, proteasome-mediated turnover of unliganded and liganded ER α on responsive promoters is an integral feature of estrogen signaling. *Mol. Cell* **11**, 695–707 (2003).
2. Lipford, J. R. *et al.* A putative stimulatory role for activator turnover in gene expression. *Nature* **438**, 113–116 (2005).
3. Yao, J. *et al.* Dynamics of heat shock factor association with native gene loci in living cells. *Nature* **442**, 1050–1053 (2006).
4. Yu, P. & Kodadek, T. Dynamics of the hypoxia-inducible factor-1-vascular endothelial growth factor promoter complex. *J. Biol. Chem.* **282**, 35035–35045 (2007).
5. Nalley, K., Johnston, S. A. & Kodadek, T. Proteolytic turnover of the Gal4 transcription factor is not required for function *in vivo*. *Nature* **442**, 1054–1057 (2006).
6. Collins, G. A., Lipford, J. R., Deshaies, R. J. & Tansey, W. P. Gal4 turnover and transcription activation. *Nature* **461**, doi:10.1038/nature08406 (2009).

doi:10.1038/nature08407